

# **Book of Abstracts**

**5<sup>th</sup> Stochastic Modeling Techniques and  
Data Analysis International Conference with  
Demographics Workshop**

# **SMTDA2018**

*Editor*

**Christos H Skiadas**

**12 - 16 June 2018**

**Cultural Centre of Chania, Crete, Greece**



**Imprint**

**Proceedings of the 5<sup>th</sup> Stochastic Modeling Techniques and Data  
Analysis International Conference with Demographics Workshop  
Chania, Crete, Greece: 12-15 June, 2018**

Published by: ISAST: International Society for the Advancement of  
Science and Technology.

Editor: Christos H Skiadas

Book ISBN: 978-618-5180-27-0

e-Book ISBN: 978-618-5180-29-4

© Copyright 2018 by ISAST: International Society for the Advancement of  
Science and Technology.

*All rights reserved. No part of this publication may be reproduced, stored,  
retrieved or transmitted, in any form or by any means, without the written  
permission of the publisher, nor be otherwise circulated in any form of binding  
or cover.*

## Preface

It is our pleasure to welcome the guests, participants and contributors to the International Conference (SMTDA 2018) on Stochastic Modeling Techniques and Data Analysis and (DEMOGRAPHICS2018) Demographic Analysis and Research Workshop.

The main goal of the conference is to promote new methods and techniques for analyzing data, in fields like stochastic modeling, optimization techniques, statistical methods and inference, data mining and knowledge systems, computing-aided decision supports, neural networks, chaotic data analysis, demography and life table data analysis.

SMTDA Conference and DEMOGRAPHICS Workshop aim at bringing together people from both stochastic, data analysis and demography areas. Special attention is given to applications or to new theoretical results having potential of solving real life problems.

SMTDA 2018 and DEMOGRAPHICS 2018 focus in expanding the development of the theories, the methods and the empirical data and computer techniques, and the best theoretical achievements of the Stochastic Modeling Techniques and Data Analysis field, bringing together various working groups for exchanging views and reporting research findings.

We thank all the contributors to the success of these events and especially the authors of this Book of Abstracts. Special thanks to the Plenary, Keynote and Invited Speakers, the Session Organisers, the Scientific Committee, the ISAST Committee, Yiannis Dimotikalis, the Conference Secretary Mary Karadima and all the members of the Secretariat.

May 2018

Christos H. Skiadas  
Conference Co-Chair

## SCIENTIFIC COMMITTEE

Jacques Janssen, Honorary Professor of Universite' Libre de Bruxelles, Honorary Chair  
 Alejandro Aguirre, El Colegio de México, México  
 Alexander Andronov, Transport and Telecom. Institute, Riga, Latvia  
 Vladimir Anisimov, Statistical Consultant & Honorary Professor, University of Glasgow, UK  
 Dimitrios Antzoulakos, University of Piraeus, Greece  
 Soren Asmussen, University of Aarhus, Denmark  
 Robert G. Aykroyd, University of Leeds, UK  
 Narayanaswamy Balakrishnan, McMaster University, Canada  
 Helena Bacelar-Nicolau, University of Lisbon, Portugal  
 Paolo Baldi, University of Rome "Tor Vergata", Italy  
 Vlad Stefan Barbu, University of Rouen, France  
 S. Bersimis, University of Piraeus, Greece  
 Henry W. Block, Department of Statistics, University of Pittsburgh, USA  
 James R. Bozeman, American University of Malta, Malta  
 Mark Brown, Department of Statistics, Columbia University, New York, NY  
 Ekaterina Bulinskaya, Moscow State University, Russia  
 Jorge Caiado, Centre Appl. Math., Econ., Techn. Univ. of Lisbon, Portugal  
 Enrico Canuto, Dipart. di Automatica e Informatica, Politec. di Torino, Italy  
 Mark Anthony Caruana, University of Malta, Valletta, Malta  
 Erhan Çinlar, Princeton University, USA  
 Maria Mercè Claramunt, Barcelona University, Spain  
 Marco Dall'Aglia, LUISS Rome, Italy  
 Guglielmo D'Amico, University of Chieti and Pescara, Italy  
 Pierre Devolder, Université Catholique de Louvaine, Belgium  
 Giuseppe Di Biase, University of Chieti and Pescara, Italy  
 Yiannis Dimotikalis, Technological Educational Institute of Crete, Greece  
 Robert Elliott, Haskayne School of Business, University of Calgary, and University of South  
 Australia, Adelaide, Australia  
 Dimitris Emiris, University of Piraeus, Greece  
 N. Farmakis, Aristotle University of Thessaloniki, Greece  
 Lidia Z. Filus, Dept. of Mathematics, Northeastern Illinois University, USA  
 Jerzy K. Filus, Dept. of Math. and Computer Science, Oakton Community College, USA  
 •Ilia Frenkel, Sami Shamoon College of Engineering, Israel  
 Leonid Gavrilov, Center on Aging, NORC at the University of Chicago, USA  
 Natalia Gavrilova, Center on Aging, NORC at the University of Chicago, USA  
 A. Giovanis, Technological Educational Institute of Athens, Greece  
 Valerie Girardin, Université de Caen Basse Normandie, France  
 Joseph Glaz, University of Connecticut, USA  
 Maria Ivette Gomes, Lisbon University and CEAUL, Lisboa, Portugal  
 Gerard Govaert, Université de Technologie de Compiègne, France  
 Alain Guenoche, University of Marseille, France  
 Y. Guermeur, LORIA-CNRS, France  
 Montserrat Guillen University of Barcelona, Spain  
 Steven Haberman, Cass Business School, City University, London, UK  
 Diem Ho, IBM Company  
 Emilia Di Lorenzo, University of Naples, Italy  
 Aglaia Kalamatianou, Panteion Univ. of Political Sciences, Athens, Greece  
 Udo Kamps, Inst. für Stat. und Wirtschaftsmath., RWTH Aachen, Germany  
 Alex Karagrigoriou, Department of Mathematics, University of the Aegean, Greece

A. Katsirikou, University of Piraeus, Greece  
 Włodzimierz Klonowski, Lab. Biosign. An. Fund., Polish Acad of Sci, Poland  
 A. Kohatsu-Higa, Osaka University, Osaka, Japan  
 Tõnu Kollo, Institute of Mathematical Statistics, Tartu, Estonia  
 Krzysztof Kolowrocki, Depart. of Math., Gdynia Maritime Univ., Poland  
 Dimitrios G. Konstantinides, Dept. Stat. & Act. Sci., Univ. Aegean, Greece  
 Volodymyr Koroliuk, University of Kiev, Ukraine  
 Markos Koutras, University of Piraeus, Greece  
 Raman Kumar Agrawalla, Tata Consultancy Services, India  
 Yury A. Kutoyants, Lab. de Statistique et Processus, du Maine University, Le Mans, France  
 Stéphane Lallich, University of Lyon, France  
 Ludovic Lebart, CNRS and Telecom France  
 Claude Lefevre, Université Libre de Bruxelles, Belgium  
 Mei-Ling Ting Lee, University of Maryland, USA  
 Philippe Lenca, Telecom Bretagne, France  
 Nikolaos Limnios, Université de Technologie de Compiègne, France  
 Bo H. Lindqvist, Norwegian Institute of Technology, Norway  
 Brunero Liseo, University of Rome, Italy  
 Fabio Maccheroni, Università Bocconi, Italy  
 Claudio Macci, University of Rome "Tor Vergata", Italy  
 P. Mahanti, Dept. of Comp. Sci. and Appl. Statistics, Univ. of New Brunswick, Canada  
 Raimondo Manca, University of Rome "La Sapienza", Italy  
 Domenico Marinucci, University of Rome "Tor Vergata", Italy  
 Laszlo Markus, Eötvös Loránd University – Budapest, Hungary  
 Sally McClean, University of Ulster  
 Gilbert MacKenzie, University of Limerick, Ireland  
 Terry Mills, Bendigo La Trobe University and Deakin University, Australia  
 Leda Minkova, Dept. of Prob., Oper. Res. and Stat. Univ. of Sofia, Bulgaria  
 Ilya Molchanov, University of Berne, Switzerland  
 Karl Mosler, University of Koeln, Germany  
 Amílcar Oliveira, UAb-Open University in Lisbon, Dept. of Sciences and Technology and  
 CEAUL-University of Lisbon, Portugal  
 Teresa A Oliveira, UAb-Open University in Lisbon, Dept. of Sciences and Technology and  
 CEAUL-University of Lisbon, Portugal  
 Annamaria Olivieri, University of Parma, Italy  
 Enzo Orsingher, University of Rome "La Sapienza", Italy  
 T. Papaioannou, Universities of Piraeus and Ioannina, Greece  
 Valentin Patilea, ENSAI, France  
 Mauro Piccioni, University of Rome "La Sapienza", Italy  
 Ermanno Pitacco, University of Trieste, Italy  
 Flavio Pressacco University of Udine, Italy  
 Pere Puig, Dept of Math., Group of Math. Stat., Universitat Autònoma de Barcelona, Spain  
 Yosi Rinott, The Hebrew University of Jerusalem, Israel  
 Jean-Marie Robine, Head of the res. team Biodemography of Longevity and Vitality,  
 INSERM U710, Montpellier, France  
 Fabrizio Ruggeri, CNR IMATI, Milano, Italy  
 Leonidas Sakalauskas, Inst. of Math. and Informatics, Vilnius, Lithuania  
 Werner Sandmann, Dept. of Math., Clausthal Univ. of Tech., Germany  
 Gilbert Saporta, Conservatoire National des Arts et Métiers, Paris, France  
 W. Sandmann, Dept. of Mathematics, Clausthal University of Technology, Germany  
 Lino Sant, University of Malta, Valletta, Malta  
 José M. Sarabia, Department of Economics, University of Cantabria, Spain

Sergio Scarlatti, University of Rome "Tor Vergata", Italy  
Hanspeter Schmidli, University of Cologne, Germany  
Dmitrii Silvestrov, University of Stockholm, Sweden  
P. Sirirangsi, Chulalongkorn University, Thailand  
Christos H. Skiadas, Technical University of Crete, Greece (Co-Chair)  
Charilaos Skiadas, Hanover College, Indiana, USA  
Dimitrios Sotiropoulos, Techn. Univ. of Crete, Chania, Greece  
Fabio Spizzichino, University of Rome "La Sapienza", Italy  
Gabriele Stabile, University of Rome "La Sapienza", Italy  
Valeri Stefanov, The University of Western Australia  
Anatoly Swishchuk, University of Calgary, Canada  
R. Szekli, University of Wroclaw, Poland  
T. Takine, Osaka University, Japan  
Andrea Tancredi, University of Rome "La Sapienza", Italy  
P. Taylor, University of Melbourne, Australia  
Cleon Tsimbos, University of Piraeus, Greece  
Mariano Valderrama, University of Granada, Spain  
Panos Vassiliou, Department of Statistical Sciences, University College London, UK  
Beatrice Venturi, University of Cagliari, Italy  
Larry Wasserman, Carnegie Mellon University, USA  
Wolfgang Wefelmeyer, Math. Institute, University of Cologne, Germany  
Shelly Zacks, Binghamton University, State University of New York, USA  
Vladimir Zaiats, Universitat de Vic, Spain  
K. Zografos, Department of Mathematics, University of Ioannina, Greece

## Plenary/Keynote Talks

In celebration of Raimondo Manca's 70th birthday and in honour of his contributions to Applied Statistics, Stochastic Modeling and Data Analysis and his support to ASMDA and SMTDA activities

### **Raimondo Manca**

Department of Methods and Models for Economics, Territory and Finance  
University of Rome "La Sapienza"  
Rome, Italy

### **N. Balakrishnan**

Department of Mathematics and Statistics  
McMaster University  
Hamilton, Ontario  
Canada

### **Mark Brown**

Department of Statistics  
Columbia University  
USA

### **Robert J. Elliott**

Haskayne School of Business,  
University of Calgary, Canada and  
Centre for Applied Financial Studies,  
University of South Australia, Adelaide, Australia

### **Olympia Hadjiliadis**

Department of Mathematics and Statistics at  
Hunter College of the City University of New York, USA

### **Markos Koutras**

Department of Statistics and Insurance Science, University of Piraeus, Greece

### **Nikolaos Limnios**

Sorbonne University, Université de technologie de Compiègne  
Applied Mathematics Lab of Compiègne (LMAC), France

### **Erol Pekoz**

Center for Information & Systems Engineering  
Boston University, USA

### **Jean-Marie Robine**

MMDN, INSERM U1198, EPHE/PSL & University of Montpellier,  
France

**Dmitrii Silvestrov**

Department of Mathematics  
Stockholm University, Stockholm, Sweden

**Nozer Singpurwalla**

Department of Systems Engineering and Engineering Management and  
Department of Management Science City University of Hong Kong and  
George Washington University, Washington, D.C., USA

**Anatoliy Swishchuk**

Department of Mathematics and Statistics  
University of Calgary, Canada

**P.-C.G. VASSILIOU**

Department of Statistical Sciences, University College London, UK,  
Mathematics Department AUTH, Greece

**Elena Yarovaya**

Lomonosov Moscow State University  
Faculty of Mechanics and Mathematics, Department of Probability Theory,  
Moscow, Russia



<b>Contents</b>	<b>Page</b>
Preface	iii
Scientific Committee	iv
Plenary/Keynote Talks	vii
Papers	1
Author Index	119



# **BOOK OF ABSTRACTS**

## **SMTDA2018**

### **Stochastic Modeling Techniques and Data Analysis International Conference with Demographics Workshop**

*In celebration of Raimondo Manca's 70th birthday and in honour of his contributions to Applied Statistics, Stochastic Modeling and Data Analysis and his support to ASMDA and SMTDA activities*

#### **Biographical and Introductory Notes by Guglielmo D'Amico**

Raimondo Manca is Full Professor in Mathematical Methods for Economics, Finance and Actuarial Sciences at the University "La Sapienza" of Rome. He studied at the University "Federico II" of Naples, Italy, where he graduated summa cum laude in Economics in 1977. He served the University "Federico II" of Naples for ten years, initially as a Research Assistant (1977 - 1981) and successively as Research Associate (1981 - 1987). From 1987 to 1994 Raimondo Manca was Associate Professor at the University "La Sapienza" of Rome before being appointed Full Professor at the University of L'Aquila in 1994. From 1995 to 1999 he worked at the University of Chieti-Pescara and then came back to Rome (University "La Sapienza") in 2000.

He started his scientific career working on Multidimensional Linear Algebra with applications to input-output models in economics. At the beginning of the eighties he started to work on semi-Markov processes and wrote his first article on homogeneous discrete time semi-Markov processes in 1981. Subsequently, he published other results with an algorithmic approach to non-homogeneous semi-Markov processes in 1984. From that moment onwards, Raimondo's research activities focused mainly on the field of semi-Markov processes with particular emphasis on applications in actuarial sciences, quantitative risk management, financial mathematics and reliability theory. He has published 75 scientific articles in peer-reviewed journals and he has written 11 scientific books: nine in English (2 Springer books and 7 ISTE WILEY) and 2 in French for Hermès Lavoisier. He is associate editor of Methodology and Computing in Applied Probability and was guest editor/co-editor for two special issues of Methodology and Computing in Applied Probability and one of Communication in Statistics Theory and Methods. As of 2018, Google Scholar reports over 1,300 citations to his work and Research Gate counts over 2,600 reads of his contributions.

## Honorary Talk

### Construction of Mean Salary Lines by Generalization of Binomial Stochastic Process

**Raimondo Manca**

*MEMOTEF, University of Roma "La Sapienza", Italy*

The salary line forecasting is an important issue for the management of the financial equilibrium of an industry and for the future revenues of its employees and their pensions. It is one of the aspect of man-power planning management. The talk will present an introduction to mean salary lines construction showing how to manage the promotions that could be automatic or by choice. The tools that will be used are a generalization of the Bernoulli stochastic processes. Some properties and some definitions of particular kind of graphs will be defined that will permit the construction of binary graphs by the recombination of binary trees.

## Plenary and Keynote Talks

### First Passage Time of Degradation Process

**Narayanaswamy Balakrishnan**

*McMaster University, Hamilton, Ontario, Canada*

In this talk, I will begin by introducing some popular degradation processes, such as Wiener process and gamma process. After presenting their basic properties, I will discuss the derivation of first passage time distribution and its approximation. I will finally present some numerical results to illustrate the methods developed.

### Taylor's Law via Ratios, for Some Distributions with Infinite Mean

**Mark Brown**

*Dept of Statistics, Columbia University, USA*

Taylor's law (TL) originated as an empirical pattern in ecology. In many sets of samples of population density, the variance of each sample was approximately proportional to a power of the mean of that sample. In a family of nonnegative random variables, TL asserts that the population variance is proportional to a power of the population mean. TL, sometimes called fluctuation scaling, holds widely in physics, ecology, finance, demography, epidemiology, and other sciences, and characterizes many classical probability distributions and stochastic processes such as branching processes and birth-and-death processes. We demonstrate analytically for the first time that a version of TL holds for a class of distributions with infinite mean. These distributions and the associated TL differ qualitatively from those of light-tailed distributions. Our results employ and contribute to methodology of Albrecher and Teugels (2006) and Albrecher, Ladoucette and Teugels (2010). This work opens a new domain of investigation for generalizations of TL. This work is joint with Professors Joel Cohen and Victor de la Pena.

## Dynamic Risk Measures and Nonlinear Expectations

**Robert J. Elliott, Samuel N. Cohen**

*University of Calgary, Canada, University of South Australia, Australia*

It is increasingly clear that a good understanding of risk and related dynamic models are required. One aspect of this is the development of simple ways of numerically representing risk. Some of these ways are well known, for example, value at risk, expected shortfall, etc. Unfortunately, these methods are static and fail to give consistent answers when considered at multiple time points. Progress has been made in developing dynamic risk measures, which give time consistent answers. Some of these ideas are related to the g-expectations considered by Peng and others. Central to the mathematical study of our consistent risk measures is the theory of Backward Stochastic Differential Equations (BSDEs). Most previous work in this area uses only noise from a Brownian motion. In our models the randomness is introduced through a martingale associated with a Markov chain. In particular we show that under certain conditions the solution to a BSDE is a time consistent risk measure.

**Keywords:** risk measure, backward stochastic differential equation, Markov chain

## Quickest Detection in The Wiener Disorder Problem with Post-Change Uncertainty

**Olympia Hadjiliadis**

*Dept of Mathematics and Statistics at Hunter College of the City University of New York, USA*

We consider the problem of quickest detection of an abrupt change when there is uncertainty about the post-change distribution. In particular, we examine this problem in the continuous-time Wiener model where the drift of observations changes from zero to a random drift with a prescribed discrete distribution. We set up the problem as a stochastic optimization in which the objective is to minimize a measure of detection delay subject to a constraint on frequency of false alarms. We design a novel composite stopping rule and prove that it is asymptotically optimal of third order under a weighted Lorden criterion for detection delay. We also analyze the conditional identification error for the post-change drift asymptotically. Our composite rules are based on CUSUM stopping times, as well as their reaction periods, namely the times between the last reset of the CUSUM statistic process and the CUSUM alarm. The established results shed new light on the performance of CUSUM strategies under model uncertainty and offer new asymptotic optimality results in this framework.

## Statistics Defined on Samples of Random Size: Theory and Applications

Markos V. Koutras

University of Piraeus, Greece

In many biological, mechanical engineering, agricultural and quality control problems some observations may get lost, for a variety of reasons, and therefore the sample size when the statistical analysis of the available data commences, should be considered as a random variable. In another setup, the sample size may depend on the occurrence of some random events, which makes the sample size random.

The probabilistic framework for these cases can be described as follows. Let  $N$  be a non-negative integer valued random variable and denote by  $X_1, X_2, \dots$  a (infinite) sequence of independent and identically distributed random variables, independent of  $N$ . Our interest focuses on the distribution of a statistic  $T=T(X_1, X_2, \dots, X_N)$  defined on a random sample  $X_1, X_2, \dots, X_N$  whose length  $N$  is a random variable. Interesting special cases of this set-up include the random sum  $T = \sum_{i=1}^N X_i$ , the random minimum  $T = \min_{1 \leq i \leq N} \{X_i\}$  and maximum  $T = \max_{1 \leq i \leq N} \{X_i\}$  or more generally the  $r$ -th smallest (largest) observation  $X_{(r:N)}$  ( $X_{(N-r+1, N)}$ ).

In the present work we provide a review of published results on the distribution of statistics  $T$  defined on samples of random size. Some interesting new results pertaining to the exact distribution of  $T$  will be provided when  $N$  belongs to wide classes of discrete distributions such as the family of power series distributions, the Panjer Family, the class of exchangeable Bernoulli mixtures and the family of Phase-Type distributions.

Finally, we illustrate how the stochastic model under study can be exploited for modeling problems arising in Financial risk management, reliability engineering, insurance science, quality control, biostatistics etc.

**Keywords:** Compound distributions; random order statistics; random sums; mixed distributions; phase-type distributions; waiting times; risk management; biostatistics; statistical quality control.

**Acknowledgment:** *The work presented in this paper has been partially funded by National Matching Funds 2016-2017 of the Greek Government, and more specifically by the General Secretariat for Research and Technology (GSRT), related to EU project "ISMPH: Inference for a Semi-Markov Process using Hazards Specification" (GA No 329128).*

## On Finite Exchangeable Sequences and their Dependence, with Actuarial Applications

**Claude Lefèvre**

*Université de Lyon 1, ISFA, LSAF EA2429, 50 Avenue Tony Garnier, F-69007  
Lyon, France*

*and*

*Université Libre de Bruxelles, Département de Mathématique, Campus de la  
Plaine C.P. 210, B-1050 Bruxelles, Belgique*

This paper deals with finite sequences of exchangeable 0-1 random variables. Our main purpose is to exhibit the dependence structure between such indicators. Working with Kendall's representation by mixture, we prove that a convex order of higher degree on the mixing variable implies a supermodular order of same degree on the indicators, and conversely. The convex order condition is then discussed for three standard distributions (binomial, hypergeometric and Stirling) in which the parameter is randomized. Distributional properties of exchangeable indicators are also revisited using an underlying Schur-constant property. Finally, two applications in insurance and credit risk illustrate some of the results.

*This is a joint work with S. Loisel (Univ. de Lyon 1) and S. Utev (Univ. of Leicester).*

**Keywords:** Exchangeable indicators, factorial series distributions, convex type orders, supermodular order, insurance risk model, credit risk portfolio.

## Branching Processes in Semi-Markov Random Environment

**N. Limnios**

*Université de Technologie de Compiègne, Sorbonne Universités, France*

We consider near critical continuous-time Markov age-dependent branching processes in semi-Markov ergodic random environment in general state space case. The law of offspring and the law of lifetimes are functions of the random environment. For this kind of processes we obtain diffusion approximation results and discuss some particular cases. An averaging result also lies to a Markov branching standard processes. In contrast to the usual transform method, (i.e., generating function approach, Laplace transform, etc.) we present a new method to obtain diffusion approximations of such processes based on semi-Markov compensating operator convergence and semimartingale relative compactness. Moreover, we prove that the near critical condition is a



necessary and sufficient condition for a diffusion approximation of a semi-Markov branching process to hold.

*Joint work with Elena Yarovaya Lomonosov Moscow State University, Russia*

**Keywords:** Branching process, semi-Markov process, diffusion approximation, near critical case, averaging.

**Mathematical Subject Classification:** 60J89, 60K15, 60J60, 60K37

#### References:

- [1]. V.S. Koroliuk & N. Limnios (2005), Stochastic systems in merging phase space, World Scientific.
- [2]. N. Limnios & E. Yarovaya (2018), A Note on the Diffusion Approximation of Branching Processes, submitted.

## Wealth Exchange Models

**Erol Peköz**

*Boston University, Boston, MA, USA*

Immediate wealth exchange models from the field of Econophysics are discrete time Markov processes where pairs of agents in a population are randomly chosen at each time step and immediately exchange wealth according to some randomized rule. The purpose of these models is to show how simple rules can yield limiting wealth distributions that arise empirically in economics. Most of these models have not been amenable to exact analyses, and the large literature on these models has typically been in the form of conducting simulations and using mean field approximations to give support for conjectures. Here we show how a family of models is amenable to a rigorous analysis yielding Dirichlet limit distributions, and we give mixing times and rates of convergence to these limits using a novel non-Markovian version of coupling from the past.

## The Return of Winter Mortality Fluctuations

**Jean-Marie Robine**

*MMDN, INSERM U1198, EPHE/PSL & University of Montpellier, France*

Excess winter mortality (EWM) vanished away in Western European countries in the 1970s. But since 2012 a return of winter mortality fluctuations is observed with a first peak of mortality observed in 2012, three years later in 2015, two years later in 2017 and one year later in 2018. The surveillance networks at first point out that this EWM was almost entirely due to the over-65s. The role of the influenza A/H3N2 virus, appeared in 1968, was suspected in 2015 then confirmed in 2017 and 2018. Over the year, the figures have been refined. In women the bulk of the excess mortality is concentrated after 75 years; 60% is explained by

the mortality of the oldest-old people (i.e., after the age of 85 years). In men, most of the excess mortality is concentrated after 65 years. The EWM of people over 65 years observed in 2015 is so strong that life expectancy at birth declines this year in almost all European countries after several decades of continuous increases. Population estimates are still missing for 2017 and 2018, but further declines in life expectancy are to be feared. The conference presents different demographic analyzes to find out what demographic mechanisms are responsible for this new EWM, namely increased mortality rates and in what proportions or ever increasing numbers of very old people exposed to these mortality rates? In the very old people, octogenarians, nonagenarians and centenarians, what is the order of magnitude, from year to year, between the size of the populations exposed and the annual number of deaths? Can we explain this return of EWM and the fluctuations which accompany it by a generalized harvesting effect. During good years, without influenza A/H3N2 outbreaks, the mortality is very low and allows the very old people to accumulate. The bad years, with an influenza epidemic dominated by the A/H3N2 virus, oldest old people are harvested blowing up the number of deaths. In subsequent years, even in poor epidemiological conditions, the number of deaths is declining as the number of oldest-old people exposed is greatly diminished. The lecture is illustrated with French data.

## Markov and Semi-Markov Reward Algorithms

**Dmitrii Silvestrov**

*Dept of Mathematics, Stockholm University, Sweden*

A survey of results on Markov and semi-Markov reward algorithms obtained in co-operation with Professor Raimondo Manca and other collaborators is presented. Effective algorithms for computing moments of rewards accumulated on trajectories of Markov chains and semi-Markov processes up to some stopping times are described. These algorithms are based on methods of sequential phase space reduction, recurrent relations connecting moments of different orders for accumulated rewards and special time-space truncation procedures. Insurance, finance, credit rating dynamics and other applications are also discussed. The survey is based on results presented in papers [1 – 6] and some latest results in the area.

**Keywords:** Markov chain, Semi-Markov process, Reward algorithm.

### References:

- [1]. F. Stenberg, R. Manca, D. Silvestrov. Semi-Markov reward models for disability insurance. *Theory of Stochastic Processes*, 12(28), 3-4, 239-254, 2006.
- [2]. F. Stenberg, R. Manca, D. Silvestrov. An algorithmic approach to discrete time non-homogeneous backward semi-Markov reward process with an application to disability insurance. *Methodology and Computing in Applied Probability*, 9, 497-519, 2007.

- [3]. E. Biffi, G. D'amico, G. Di Biase, J. Janssen, R. Manca, D. Silvestrov. Monte Carlo semi-Markov methods for credit risk migration and Basel II rules. 1 and 2. *Journal of Numerical and Applied Mathematics*, 1(96), Part 1: 28-58, Part 2: 59-86. 2008.
- [4]. D. Silvestrov, E. Silvestrova, R. Manca. Stochastically ordered models for credit rating dynamics. *Journal of Numerical and Applied Mathematics*, 1(96), 206-215. 2008.
- [5]. D. Silvestrov, R. Manca, E. Silvestrova. Computational algorithms for moments of accumulated Markov and semi-Markov rewards. *Communications in Statistics – Theory and Methods*, 43, 7, 1453-1469. 2014.
- [6]. D. Silvestrov, R. Manca. Reward algorithms for semi-Markov processes. *Methodology and Computing in Applied Probability*, DOI 10.1007/s11009-017-9559-2, 19 pp., 2017.

## Likelihood, Confirmational Tenacity, and Mood Transitions, in Bayesian Inference

**Nozer D. Singpurwalla**

*The City University of Hong Kong, Hong Kong*

Bayesian inference has now become the mainstay of statistical practice in science, engineering, and medicine, with relevance to machine learning and information fusion, being two of the current highly touted topics in data science. Simplicity and comprehensiveness aside, coherence has been a declared hallmark of Bayesian statistics. By coherence, we mean a faithful adherence to the calculus of probability. But in endeavoring to do so, a Bayesian engages in the three underlying maneuvers that remain to be made explicit; these are: i) a transition of mood from the indicative (or factual) to the irrealis (or subjunctive), and then back again to the indicative; ii) the introduction of a notion that is external to probability, namely, the likelihood; and iii) invoking a controversial behavioristic principle called “Bayesian Conditionalization”, or to the philosopher Isaac Levi, “Confirmational Tenacity”.

Mood transition is necessitated by the feature that all declared probabilities must be in the indicative mood, whereas by definition, all conditional probabilities are, in the irrealis mood. The purpose of this paper is to raise awareness to these matters, which are implicit to what Dennis Lindley has referred to as a “turning of the Bayesian crank”, and to make transparent the interplay between the Bayesian inferential mechanism and the philosophical underpinnings of probability theory. Our conclusion is that the routinely practiced Bayesian prior to posterior iteration only mimics the probability calculus, and in so doing casts pallor on the claim of coherence. Bayesian inference is therefore approximately coherent. This means that alternative paradigms with a Bayes label, such as Empirical Bayes, Objective Bayes, Pseudo Bayes, and Belief Functions, many of which are an anathema to Bayesian royalty, should have a place at the table.

**Keywords:** Bayesian Conditionalization, Coherence, Confirmation Tenacity, Subjective Likelihoods, Subjective Probability.

## **Stochastic Modelling of Big Data in Finance**

**Anatoliy Swishchuk**

*Dept of Mathematics and Statistics, University of Calgary, Canada*

Big data has now become a driver of model building and analysis in a number of areas, including finance. More than half of the markets in today's highly competitive and relentlessly fast-paced financial world now use a limit order book mechanism to facilitate trade. In this talk we show how to model and to deal with big data arising in electronic markets for algorithmic and high-frequency (milliseconds) trading that contain two types of orders, limit orders and market orders. In particular, we consider several extensions, suggested by empirical observations, of Cont and Larrard (SIAM J. Fin. Math., 2013) Markovian stochastic model. One of the extensions is to extend their framework to 1) arbitrary distributions for book events inter-arrival times (possibly non-exponential), and 2) both the nature of a new book event and its corresponding inter-arrival time depend on the nature of the previous book event (not independent). The dynamics of the bid and ask queues are modeled by Markov renewal process and the mid-prices is modelled by a semi-Markov process. The second extension is associated with the case when the price changes are not fixed at one tick. And the third one is related to the case with arbitrary number of states for the price changes. We justify and illustrate our approaches by calibrating our model to the five stocks, Amazon, Apple, Google, Intel, Microsoft, on June 21st, 2012 (Lobster data), to the 15 stocks from Deutsche Boerse Group (September 23d, 2013), and to Cisco asset (November 3d, 2014).

## **Change of Probability Measure in Non-Homogeneous Semi-Markov and Markov Renewal Processes. An Application to Forward Probability Measure in Credit Risk**

**P.-C.G. Vassiliou**

*Dept of Statistical Sciences University College London;  
Mathematics Dept, AUTH*

In the present we study changes of probability measures in non-homogeneous semi-Markov chains and Markov renewal processes. As an application the valuation of a defaultable bond is done with the use of the forward probability measure of no default up to maturity time in an inhomogeneous semi Markov and an inhomogeneous Markov renewal process. It is proved that under the forward probability measure, the semi Markov property is maintained and also the foundation of an inhomogeneous Markov renewal process. Martingale characterization theorems are proved for both inhomogeneous semi-Markov and Markov

renewal processes. These are provided in a form suitable in order to find simplified closed analytic functional relationships, between the forward probability measure transition probabilities sequences and the corresponding real world probability measure ones. We also provide a general theorem where some general necessary and sufficient conditions are given that should be satisfied when we calibrate our models. An algorithm is established for the calibration of the model in the case of the migration process in credit risk. Finally an important application is discussed in the case of the migration process of defaultable bonds in credit risk.

## **Particle Systems, Non Homogeneous Environments and Large Deviations**

**Elena Yarovaya**

*Dept of Probability Theory, Faculty of Mechanics and Mathematics,  
Lomonosov Moscow state University, Moscow, Russia*

The modeling of the evolution of particle or cell populations in physics and biology has given an important incentive to the development of stochastic processes. The achievements of recent decades in these sciences have created many open problems. To solve them, we needed to introduce new classes of stochastic processes which allow to conduct the modeling and analysis of populations with complex structures and dynamics. In particular stochastic processes that model populations distributed in space as well as their genealogies will be considered. This will include branching random walks.

Nowadays it is commonly accepted that branching random walks are crucially useful in investigations of stochastic systems with birth, death and migration of their elements. In some of these branching random walks the phase space, in which the particles live, can be naturally assumed to be continuous, in other cases this phase space may be a finite or infinite set of discrete points. In the talk, the principal attention will be paid to the properties of branching random walks on multidimensional integer lattices. We will be mainly interested in the problems related to the limiting behavior of branching random walks such as existence of phase transitions under change of various parameters, the properties of the limiting distribution of the particle population, existence and the shape of the so-called propagating fronts of particles, etc. Naturally, the answer to these and other questions heavily depend on numerous factors which affect the properties of a branching random walk. Therefore, we will try to describe, rather detailed, how the properties of a branching walk depend on the fact of randomness (or determinacy) of the branching media and its homogeneity, on the number and mutual disposition of the branching sources, and also on such properties of a branching walk as its symmetry

and finiteness or infiniteness of the variance of jumps. Some spatio-temporal evolutionary models of particle population will be discussed. We present also some results of simulation of branching random walks and discuss how they may be applied to numerical estimation on bounded time intervals, of various characteristics describing the properties of the phase transitions.

**Key Words:** branching random walks, spatial structure, heavy-tailed distributions, phase transitions

**Acknowledgments:** *The work was supported by the Russian Foundation for Basic Research, project no. 17-01-00468.*

## Invited and Contributed Talks

### Stochastic Modeling and Data Analysis of Arab Impact Factor

**Mahmoud Abdel-Aty**

*Applied Science University, kingdom of Bahrain*

Classifying researchers according to the quality of their published work rather than the quantity is a curtail issue. We discuss the general classification of some journals and how one creates sufficient database to be used for an annual report production. Also, we introduce a new formula to test the researcher's production. The suggested equation depends on the number of the single-author published papers and their citations to be added as a new factor to the known h-index. These factors give an advantage and make a clear evidence of innovative authors and reduce the known h-index for authors who are gaining citations by adding their names to multi-author papers. It is shown that various dimensions of ethical integrity and originality will be effective in this new index. An important scenario arising from the analysis is shown in terms of examples. It refers to larger differences between the h- and the new index which comes from the whole work and the one comes from the single-author papers only, is shown [1,2].

#### References:

- [1]. M. Abdel-Aty, New Index for Quantifying an Individual's Scientific Research Output, arXiv:1305.6026
- [2]. M. Abdel-Aty, Indices to Quantify the Ranking of Arabic Journals and Research Output, Inf. Sci. Lett. 1, (2012)

### Selection of proximity measures for a Topological Correspondence Analysis

**Rafik Abdesselam**

*COACTIS-ISH Management Sciences Laboratory - Human Sciences Institute,  
University of Lyon, Lumiere Lyon 2, France*

In this paper, we propose a new approach which compare and classify proximity measures in a topological structure to select the "best" one for a Topological Correspondence Analysis. Measures of proximity play an important role in many domains of data analysis. The results of any investigation whether association exists between variables or any operation of clustering or classification of objects are strongly depend on

the proximity measure chosen. The user has to select one measure among many existing ones. Yet, according to the notion of topological equivalence chosen, some measures are more or less equivalent. The concept of topological equivalence uses the basic notion of local neighborhood. We define the topological equivalence between two proximity measures, in the context of association between two qualitative variables, through the topological structure induced by each measure.

We compare proximity measures and propose a topological criterion for choosing the "best" association measure, adapted to the data considered, among some of the most used proximity measures for qualitative data. The principle of the proposed approach is illustrated using a real data set with conventional proximity measures of literature for qualitative variables.

**Keywords:** Proximity measure; topological structure; neighborhood graph; adjacency matrix; topological equivalence; topological index of independence.

## Updating of PageRank in Evolving Tree Graphs

**Benard Abola<sup>1</sup>, Pitso Selekwa Biganda<sup>1,2</sup>, Christopher Engström<sup>1</sup>,  
Godwin A. Kakuba<sup>3</sup>, John M. Mango<sup>3</sup> Sergei Silvestrov<sup>1</sup>**

<sup>1</sup>Division of Applied Mathematics, The School of Education, Culture and Communication, Mälardalen University, Sweden, <sup>2</sup>Dept of Mathematics, College of Natural and Applied Sciences, University of Dar es Salaam, Tanzania, <sup>3</sup>Dept of Mathematics, School of Physical Sciences, Makerere University, Uganda

Updating PageRank refers the process of computing new PageRank values after change(s) has occurred in a graph. The main goal of the updating is to avoid re-calculating the values from scratch. It is known from literature that handling PageRank's update is problematic, in particular when it involves both link and page updates. In this talk we will focus on updating PageRank of evolving tree graphs. We will describe how to maintain level structures when a cycle has been created, and also investigate the practical and theoretical efficiency to update PageRanks for an evolving graph with a single cycle. In all these cases, we will consider PageRank as the expected number of visits to target vertex if a multiple random walks is performed, starting at each vertex once and weighting each of these walks by weight value.

**Keywords:** PageRank, Random walk, Graph, Link-update, Page-update.



## HAR-type models for realized volatility forecasting: the role of structural breaks

Giuseppina Albano<sup>1</sup>, Davide De Gaetano<sup>2</sup>

<sup>1</sup>Dept of Statistics and Economics, University of Salerno, Italy

<sup>2</sup>SOSE - Soluzioni per il Sistema Economico S.p.A., Italy

The paper addresses the problem of forecasting realized volatility in the context of HAR-type models when structural breaks are present. In particular we consider several models relying on the decomposition of realized variance in its positive or negative (semivariance) part and its continuous or discontinuous part (jumps).

In order to handle parameter instability, some specific forecast combinations, based on different estimation windows with alternative weighting schemes, are discussed. The forecasting performance of the proposed forecast combinations are compared in terms of suitable loss functions, by using the Model Confidence Set (MCS) procedure, on several real Datasets.

**Keywords:** Realized volatility, HAR-type models, Forecast combinations.

## New Distributions to Mortality Modelling: A Different Approach

Andreopoulos Panagiotis<sup>1</sup>, Tragaki Alexandra<sup>1</sup>,  
Bersimis G. Fragkiskos<sup>2</sup>

<sup>1</sup>Dept of Geography, Harokopio University, Greece, <sup>2</sup>Dept of Informatics and Telematics, Harokopio University, Greece

This paper presents the comparison's results between traditional mortality models (Gompertz Makeham) versus innovative mortality models such as B.ANBE & ANBE.G, as regards their distribution parameters. This study illustrates the population's mortality of all ages in Greece in the 21st century. The proposed B.ANBE distribution is a six parameter distribution known as the Beta Gompertz Generalized Makeham distribution which is quite flexible and presents a decreasing, increasing and bathtub-shaped failure rate function depending on its parameters making it more effective in modelling mortality data and reliability problems. In addition, the new ANBE.G distribution is a five parameter distribution known as the Gamma Gompertz Generalized Makeham distribution which is also flexible for different data sets. In order to show the distributions' flexibility, an application using a real data set is presented, with some important statistical properties of the aforementioned distributions.

**Keywords:** B.ANBE distribution, ANBE.G distribution, Gompertz Makeham distribution, Mortality

## Methods for Assessing Critical States of Complex Systems

**Valery Antonov**

*Peter the Great St.-Petersburg Polytechnic University, Russia*

In the study of the body as a complex system, a large number of characteristics are required to assess the current state and forecast their development. Often this measurement is impossible due to lack of time and equipment. Therefore, it becomes necessary to assess the state and short-term dynamics of a change in the system, using characteristic signals that can be obtained in real time. It is assumed that the system measures the periodic character RT, which is a discrete sequence of intervals  $r_i$ . Under these conditions, the investigation of complex system is reduced to analyzing the time series. This paper presents an overview existing research methods. Statistical methods are used to directly quantify the signal in the time interval under study (dispersion, coefficient of variation, etc.) Time analysis consists in studying the law of intervals distribution as random variables. Autocorrelation analysis makes it possible to recognize the latent periodicity in RT. Frequency analysis of the oscillations spectral power provides information on the power distribution as a function of the oscillations frequency. Correlation rhythmography consists in graphical representation of consecutive pairs of intervals (previous and next) in a two-dimensional coordinate plane. Fractal analysis is assessing fractal characteristics of data. Mmethod WTMM (wavelet transform modulus maxima) uses plotting of the local maximum line of wavelet transform. Holder exponent, a measure of the degree to which a signal is differentiable, is used to detect the presence of damage and when that damage occurred.

**Keywords:** complex system, critical state, time series, research methods

## Side Effects of New Drug Computer-aided Design by Methods of Quantitative Structure-Activity Relationship

**Valery Antonov<sup>1</sup>, Anatoly Kovalenko<sup>2</sup>, Konstantin Lebedinskii<sup>3</sup>**

*<sup>1</sup>Dept of Mathematics, Peter the Great St.Petersburg Polytechnic University, Russia, <sup>2</sup>Ioffe Institute, Russian Academy of Sciences, Russia, <sup>3</sup>Dept of Anaesthesiology and Reanimation, St. Petersburg Medical Academy of Postgraduate Studies, Russia*

We discussed the procedure to build models which enable to prognosticate their various properties based on the structure and formulation of chemical compounds. We analyzed the mathematical descriptors which represent numerically the characteristics of ligands and bio-targets, as well as their complexes intended to achieve the

pharmacological effect required in the research trend selected, with regard to the problems in computer-aided design of new drug products. We drew attention to the side effects as a number of other types of biological activities inherent in the substance under study and conditioned by the opposite effect of the drug product used in various dosages, and even by small deviations in the formulation and concentrations of its components. We described a method to prognosticate the potential spectrum of biological activity of compounds, based on the classical Bayesian approach to assessing the probability of these effects.

**Keywords:** Quantitative analysis, descriptors, side effects, prognosis of probable spectrum

## **Support Vector Machines: A Review and Applications in Statistical Process Monitoring**

**Anastasios Apsemidis and Stelios Psarakis**

*Athens University of Economics and Business, Dept of Statistics*

A few years ago, Statistical Process Monitoring and Statistical Learning began to interact with each other, in order to solve difficult and complex industrial problems. In this paper, we examine a specific statistical learning technique named Support Vector Machines. It is a most powerful algorithm, although coming from a really simple idea, which is increasingly used in many fields of statistics and computer science. We present a thorough review of the literature concerning support vector machines in the process monitoring field, we test one of the mentioned works on a real data set and try to improve its performance using Principal Components Analysis. Finally, we present an alternative approach which is able to yield better results.

## **I1-Penalized Censored Gaussian Graphical Model**

**Luigi Augugliaro, Antonino Abbruzzo, Veronica Vinciotti**

*University of Palermo, Italy*

Graphical lasso is one of the most used estimators for inferring genetic networks. Despite its diffusion, there are several fields in applied research where the limits of detection of modern measurement technologies make the use of this estimator theoretically unfounded, even when the assumption of a multivariate Gaussian distribution is satisfied. Typical examples are data generated by polymerase chain reactions and flow cytometer. The combination of censoring and high-dimensionality make inference of the underlying genetic networks from these data very

challenging. In this paper we propose an l1-penalized Gaussian graphical model for censored data and derive two EM-like algorithms for inference. By an extensive simulation study, we evaluate the computational efficiency of the proposed algorithms and show that our proposal overcomes existing competitors when censored data are available. We apply the proposed method to gene expression data coming from microfluidic RT-qPCR technology in order to make inference on the regulatory mechanisms of blood development.

**Keywords:** Censored data; Expectation-Maximization algorithm; Gaussian graphical model; Graphical Lasso; High-dimensional Data.

## A Proposed Procedure for Forecast Future Mortality Trends

Avraam Demetris<sup>1</sup>, Bersimis. G. Fragkiskos<sup>2</sup>,  
Andreopoulos Panagiotis<sup>3</sup>

<sup>1</sup>Dept of Business and Public Administration, University of Cyprus, Cyprus, <sup>2</sup>Dept of Informatics and Telematics, Harokopio University, Greece, <sup>3</sup>Dept of Geography, Harokopio University, Greece

Many mathematical models have been developed to analyse and forecast human mortality dynamics. Some of the models are composite in structure but able to reproduce the entire age-specific mortality patterns, including the observed peculiarities at young and extremely old ages, while other models are simple in structure but able to reproduce only specific parts of those patterns. Although the simple models can be easily applied with most of the statistical fitting procedures, the composite models are often preferable as they provide higher accuracy on fitting and forecasting actual death rates. In this study, we propose a new procedure for forecasting mortality trends where the Gompertz function is used in a way to give highly accurate predictions. In this method, we fit the Gompertz model to death rates of consecutive periods and we extrapolate the trends of the estimated model parameters. We also observe that the residuals between the fitted and the observed values, obtained by fitting period death rates with the Gompertz model, generate approximately stationary profiles. We therefore use the extrapolated Gompertz parameters to predict future mortality trajectories and we add the stationary profile of the residuals to these trajectories. We demonstrate the procedure with a back-test application to Swedish and Greek death rates and we compare the results with Lee-Carter predictions.

**Keywords:** Gompertz, Lee-Carter, mortality rate, forecast, compensation effect, Strehler-Mildvan correlation, back-test analysis

## Differential Gene Expression and Point Pattern Analysis

**Guillermo Ayala<sup>1</sup>, Angela Riffo-Campos<sup>1</sup> and Amelia Simó<sup>2</sup>**

<sup>1</sup> *Departamento de Estadística e Investigación Operativa. Universidad de Valencia. Spain*

<sup>2</sup> *Departamento de Matemáticas - IMAC. Universitat Jaume I, Castellón, Spain*

This paper is concerned with gene set differential expression analysis using statistical analysis of point patterns. Given an expression matrix, a phenotypic binary covariable and a given gene set (subset of rows in the expression matrix) different point patterns can be considered. Self-contained null hypothesis usually considered in differential expression analysis can be formulated in terms of these associated point processes. We focus in the particular case of two paired groups of samples; in this case, spatial point process methods to test the possible raised density of points near a fixed point can be used. Two classes of methods are proposed: method based on distances and methods based on likelihood. This approach will alleviate the well-known problem of the low sample sizes with respect to the large number of genes studied.

**Key Words:** Gene set differential expression; RNAseq; Point processes; non-homogeneous Poisson point process; Randomi labelling.

## Modelling Customer Behaviour In Online Gambling

**Maria Balzan, Lino Sant**

*Dept of Statistics and Operations Research,  
Faculty of Science, University of Malta, Malta*

In this paper a continuous Markov chain is constructed to model the behavior of gaming customers while they engage online alternating between periods of inactivity and high intensity interaction. An extensive database containing daily transactions provided by a major operator was analyzed statistically. Using various clustering methods a suitable state-space capturing the categorization of customers was defined. This enabled Markov chain models to be fully specified for various categories of customers. Corresponding transitions probabilities and survival times were estimated.

**Keywords:** Online gambling modelling, Markov chains, customer categorization.

## Life expectancy in India- Present and Future trends A Spatio-Temporal Bayesian Model Approach

Adrita Banerjee, Ajeet Kumar Singh  
IIPS, Mumbai, India

### Background

Global Aging poses global challenge mostly in terms of large effect on size and shape of government budgets. To plan for health and social services and pensions, forecasts of future mortality and life expectancy are needed (Bennett et al., 2015). India is moving towards the nomenclature of 'aging population' with 8% of elderly population (Census, 2011). With fast increase in elderly population there is a burden on the government budget for meeting the health care and social and economic needs of the elderly. Thus prediction of future trend in life expectancy and inequality in life expectancy is essential for planning purpose. There are no studies in India which have dealt in estimating the life expectancy at the state level and has examined the individual based measurement of inequality in life expectancy. The present paper tries to forecast the life expectancy for country and the states using Sample Registration Data.

### Data Source and Method

Using data from the Sample Registration System (SRS) for the period 1981–2011 the trend in life expectancy in India and major states and life expectancy till 2040 has been projected. Bayesian spatiotemporal models have been used for forecasting of age-specific mortality and life expectancy at national and state-level. The models included components that accounted for mortality in relation to age, and space i.e. state. In order to test the model performance the projected life expectancy provided by United Nations for the country as a whole has been matched with the projected findings.

### Findings

Life expectancy in India has substantially increased for the last three decades. For men, it has risen from 51 years in 1981 to 66 years in 2011, and for women, it has increased from 55 years in 1981 to 71 years in 2011. National life expectancy in 2040 is expected to reach to 75.9 years for females, 70.7 year for males and overall it is projected as 73.1 years. The states also show a similar trend.

### Conclusion

The life expectancy has been on a rise and is expected to rise further which implies a sufficient part of the health budget should be allocated for elderly population (Geriatric Care). Health and social policies are needed to curb widening life expectancy inequalities, help lag off states to catch up in the longevity gains.

## On the optimal designs for the prediction of complex Ornstein-Uhlenbeck processes

Sándor Baran, Kinga Sikolya

*Faculty of Informatics, University of Debrecen, Hungary*

Complex Ornstein-Uhlenbeck (OU) processes have various applications in statistical modelling. They are playing roles e.g. in the description of the motion of a charged test particle in a constant magnetic field (Balescu, 1997) or in the study of rotating waves in time-dependent reaction diffusion systems (Beyn and Lorenz, 2008; Otten, 2015), whereas Kolmogorov used such a process to model the so-called Chandler wobble, the small deviation in the Earth's axis of rotation (Lambeck, 1980). In these applications parameter estimation and model fitting is based on discrete observations of the underlying stochastic process, however the accuracy of the results strongly depends on the observation points. We study the optimal design problem for prediction of a complex OU process on a monotonic set with respect to integrated mean square prediction error (IMSPE) and entropy criterion. We derive the exact form of IMSPE and entropy criterion, moreover, we show that the optimal designs based on entropy criterion are equidistant. Finally, we present some numerical experiments to illustrate selected cases of optimal designs for small number of sampling locations.

**Keywords:** Chandler wobble, complex Ornstein-Uhlenbeck process, integrated mean square prediction error, entropy

**Acknowledgments.** *This research has been supported by the ÚNKP-17-4 New National Excellence Program of the Ministry of Human Capacities project.*

### References

- [1]. Balescu, R. Statistical Dynamics- Matter out of Equilibrium. Imperial College Press, London, 1997.
- [2]. Bayn, W. J. and Lorenz, J. Nonlinear stability of rotating patterns. *Dyn. Partial Differ. Equ.* 5, 349--400, 2008.
- [3]. Lambeck, K. The Earth's Variable Rotation: Geophysical Causes and Consequences. Cambridge University Press, Cambridge, 1980.
- [4]. Otten, D. Exponentially weighted resolvent estimates for complex Ornstein-Uhlenbeck systems. *J. Evol. Eq.* 15, 753--799, 2015.

## Statistical Inference for Multi-State Systems based on semi-Markov processes

Vlad Stefan Barbu<sup>1</sup>, Alex Karagrigoriou<sup>2</sup>, Andreas Makrides<sup>1</sup>

<sup>1</sup>Université de Rouen, Laboratoire de Mathématiques Raphaël Salem Saint-Étienne-du-Rouvray, France,

<sup>2</sup>Dept of Mathematics, University of the Aegean, Greece

In this work we focus on multi state systems that we model by means of semi-Markov processes. The sojourn times are seen to be independent not identically distributed random variables and assumed to belong to a general class of distributions that includes several popular reliability distributions like the exponential, Weibull, and Pareto. We obtain maximum likelihood estimators of the parameters of interest and investigate their asymptotic properties. Plug-in type estimators are furnished for various quantities related to the system under study.

**Keywords:** Multi-state system, reliability theory, survival analysis, reliability indicators, semi-Markov processes, parameter estimation.

## Calculation Methods for Binomial Classification of Discrete Data. An Application in both Simulated and Real Data

Bersimis Fragkiskos<sup>1</sup>, Varlamis Iraklis<sup>1</sup>,  
Vamvakari Malvina<sup>1</sup>, Demosthenes Panagiotakos<sup>2</sup>

<sup>1</sup>Dept of Informatics and Telematics, Harokopio University, Greece, <sup>2</sup>Dept of Nutrition Science-Dietetics, Harokopio University, Greece

This work attempts to evaluate classification methods of binomial characteristics based on discrete data. The methods used are logistic regression, classification trees, neural networks and Support vector machine techniques, as well as some special modifications of them. The methods were applied in both simulated and actual medical data. A variety of scenarios was applied based on the distribution's parameters of the component variables, as well as, based on the number of components used. All methods were assessed as to their classification performance by using area under the receiver operating characteristic curve, true positive rate, true negative rate, true rate, positive predictive value and negative predictive value. The criteria's statistical significance was evaluated by using Bootstrap confidence intervals. Predictability was assessed by using multiple training and test samples. Results indicate that for specific methods, the corresponding criteria achieved greater values compared to the others, according the data distribution or the components' multitude. These findings provide an accurate methodology for selecting more



suitable techniques for predicting the clinical status of a person in the case of general or specific populations.

**Keywords:** AUC, Sensitivity, Specificity, Bootstrap, Classification methods

## **A New Technique for Monitoring Public Health Combining Scan Statistics and Control Charts**

**Sotiris Bersimis, Athanasios Sachlas, Markos V. Koutras**

*Dept of Statistics & Insurance Science, University of Piraeus, Greece*

To timely and efficiently detect disease outbreaks researchers should take into account both spatial and temporal dimensions. Of interest are global changes in the number of new disease events on time and/or hotspots of disease events which may evolve into outbreaks. A key assumption in monitoring public health is that under normal conditions events are uniformly distributed in the plane and there are not too many changes over time. In this work, we propose a new two-step monitoring procedure with which we monitor the number of disease events by combining scan statistics and control charts.

**Keywords:** Control Charts, Monitoring, Scan Statistics

## **A Special Gen(d)re of Big Data: Potentials of the Data Revolution to Model Gender (im)Balance**

**Gaia Bertarelli<sup>1</sup>, Fulvia Mecatti<sup>2</sup>, Franca Crippa<sup>2</sup>**

*<sup>1</sup>Dept of Economy and Management, University of Pisa, Italy,*

*<sup>2</sup>University of Milano-Bicocca, Italy*

It is suggested by the extraordinary growth in the availability of large gender-related datasets and statistical information marked by the 2015 deadline of the UN Millennium Development goals, with a focus on how to supplement, complement or even replace current production of gender statistics and modelling of gender (im)balance. An engendered statistical reasoning appears motivated by the increasing demand of gender-sensitive statistical information coming from society, official agencies and the economy, gender balance being globally recognized as a crucial objective for economic growth and development for society as a whole, for both women and men.

The 2000 UN Millennium Declaration and the eight Millennium Development Goals have engaged the world on a strict agenda to systematically monitor and report country progresses, on the basis of a shared system of measurable parameters and statistical indicators, which

has provided, in the process, voluminous data of controlled quality at country level and comparable in time and space. Moreover, gender equality is included in the 2030 US Agenda of Sustainable Development Goals, both as a particular Goal (5): “To achieve gender equality and empower all women and girls”, and as fundamental to delivering on the promises of sustainability, peace and human progress. The passage from the 2015 MDGs and the 2030 SDGs appears as having pinpointed a key turning point in the availability of gender-sensitive data, with an unprecedented outbreak of good quality and easy access gender-sensitive data.

This phenomenon represents a big step in the development of a Gender Statistics and a peculiar challenge ever since its original interpretation as a mere data disaggregation between men and women. A critical review of the current data richness, accessibility and usability for statistical purposes will be given and three key points, calling for methodological innovations, will be discussed: 1) the potential of data mining and statistical learning methods to develop improved, effective and genuinely multivariate gender statistics, able to go beyond simple and composite indicators, as mainly used in this field; 2) the need of effective methodologies for selecting relevant data streams and significant variables; and 3) the emerging possibility for elaborating the “two-speed effect” usually revealed by the gender perspective, namely from denied basic rights to non-equal opportunities between men and women. It would be therefore possible to go deeper in the study of the causation and determinants of the “transition”, and of risk factors for national gender imbalance and gender-based obstacles to economic growth and human development.

**Keywords:** Gender balance, data mining, statistical learning, two-speed effect.

## **Exploring the Relationship Between Ordinary PageRank, Lazy PageRank and Random Walk with Backstep PageRank for Different Graph Structures**

**Pitos Seleka Biganda<sup>1,2</sup>, Benard Abola<sup>2</sup>, Christopher Engström<sup>2</sup>, John Mango Magero<sup>3</sup>, Godwin A. Kakuba<sup>3</sup>, Sergei Silvestrov<sup>2</sup>**

<sup>1</sup>*Dept of Mathematics, College of Natural and Applied Sciences, University of Dar es Salaam, Tanzania,* <sup>2</sup>*Division of Applied Mathematics, The School of Education, Culture and Communication (UKK), Mälardalen University, Sweden,*

<sup>3</sup>*Dept of Mathematics, School of Physical Sciences, Makerere University, Uganda*

Ordinary PageRank, introduced by Brin and Page, may be considered as the probability that an internet surfer hits a given webpage in a type of random walk. That is, being at a particular webpage, the surfer visits the

next page with uniform probability. There are two other random walks in PageRank problem, these are, lazy random walk and random walk with backstep. The corresponding PageRank is lazy PageRank and random walk with backstep PageRank, respectively. The three walks reveal different behaviour of the surfer in graph structures. However, the relationship between ordinary PageRank, lazy PageRank and random walk with backstep PageRank is not well understood. In this talk, we will show how the three variants of PageRank depend on hyperlink matrices and the damping factor  $c$ . We will also describe how the ordinary PageRank can be formulated from lazy PageRank and random walk with backstep PageRank.

**Keywords:** Graph, lazy PageRank, random walk with backstep.

## **The Challenge of Assessing Financial Literacy: When Methodology Could Make a Difference**

**Paola A. Bongini<sup>1</sup>, Paola Iannello<sup>2</sup>, Emanuela E. Rinaldi<sup>3</sup>,  
Mariangela Zenga<sup>1</sup>, Alessandro Antonietti<sup>2</sup>**

<sup>1</sup>*University of Milano-Bicocca, Italy*, <sup>2</sup>*Università Cattolica del Sacro Cuore, Italy*,

<sup>3</sup>*Università degli Studi di Udine, Italy*

The assessment of financial literacy levels is nowadays widely recognized as necessary to both design effective financial education and evaluate its actual impact. Nevertheless, financial literacy has been assessed through different methodological approaches across studies. To address the lack of a common and agreed upon methodology to measure financial literacy, in 2011 the OECD-INFE network developed a core questionnaire [1] to be administered across a wide range of countries.

The present paper focuses on the assessment approach used to evaluate financial literacy among the Italian adult population through the administration of the OECD-INFE questionnaire. The study proposes alternative methods – Item Response Theory and Classification and Regression Trees – to treat the survey data, thus showing that different methods to analyze data could lead to diverse outputs.

The study highlights the crucial role that the methodology issue plays in assessing financial literacy. Specifically, the paper suggests that financial literacy research should open up to alternative and new methodological approach in order to obtain ever more reliable measures of financial literacy, genuinely able to capture and effectively address the educational needs of different population groups.

**Keywords:** financial literacy index; financial literacy evaluation; IRT analysis; CART analysis; target of financial education

### **References**

[1] OECD INFE (2011) Measuring Financial Literacy: Questionnaire and Guidance Notes for Conducting an Internationally Comparable Survey of Financial Literacy. OECD, Paris.

## Redistricting, Including Boundary Districts, Utilizing the Convexity Ratio and Counties

**James R. Bozeman, Jillian Mori, Abigail Salvadore**  
*American University of Malta, Malta*

We redistrict many states, avoiding gerrymandering, by starting with counties and their populations. The voting districts themselves are made as 'nicely shaped' as possible, using the convexity ratio as a guide. Those districts with immovable boundaries are dealt with appropriately.

**Keywords:** redistricting, gerrymandering, convexity ratio, boundary districts

## Assessment of Residual Lifetime Based on the Gompertz–Makeham Law and its Application to Insurance in the Global Warming Environment

**Bushinskaya A.V.<sup>1,2</sup> Timashev S.A.<sup>1,2,3</sup>**

*<sup>1</sup>Science and Engineering Center "Reliability and safety of large systems and machines" Ural branch, Russian academy of sciences, Russia, <sup>2</sup>Ural Federal University, Russia, <sup>3</sup>Old Dominion University, USA*

In terms of reliability theory, the Gompertz–Makeham (G-M) mortality law is a law of failure, where the risk norm is a combination of age-related and non-related failures. The paper describes some results of applying the Gompertz–Makeham mortality law to the basic mortality – table affected by a series of random number of accidents per year of different nature that resulted in a random number of mortalities without prior predisposition to death or wear. The difference with the classic G-M law is in that the age-independent component is associated with different type of extreme situations including accidents induced by global warming.

The parameters of the G–M law for a local cohort were estimated using nonlinear regression methods. The paper describes an accessible four-point method that allows calculating estimates of the parameters of the G–M law on the basis of numbers of survivors at four equidistant ages (at the age of 20, 40, 60, and 80 years). The method has been tested on variety of specific mortality tables that reflect the influence of different environments of the Makeham-type part of the G-M law on the mortality. The accidents were modeled using the Monte-Carlo method. Results of calculations are presented and a correlation between the G-M parameters and the hypothetical climate shift is described.

In this paper the Monte-Carlo method in conjunction with the G-M model is also used for assessing the number of mortalities for a given system of interdependent critical infrastructures located in a given region and

calibrating the insurance premium using direct and indirect methods using company bankruptcy as the ultimate criteria.

## **The Toll of Incidents, Accidents and Disasters on the Average Life Expectancy in Good Health**

**Bushinskaya A.V.<sup>1,2</sup> Timashev S.A.<sup>1,2,3</sup>**

*<sup>1</sup>Science and Engineering Center "Reliability and safety of large systems and machines" Ural branch, Russian academy of sciences, Russia, <sup>2</sup>Ural Federal University, Russia, <sup>3</sup>Old Dominion University, USA*

The paper presents an algorithm for estimating the decrease of the average life expectancy (ALE) as a result of an accident, a catastrophe or a terrorist act, accompanied by human casualties or loss of limb, health. Solution of this problem is necessary when managing territorial risk by the ALE criterion.

If the probability of non-recoverable losses is zero, the safety problem transforms into a purely structural reliability problem. To use ALE as a generalized territorial criterion for risk optimization, it is necessary to correctly determine it from demographic positions. Hence, it is necessary to connect all the consequences of the loss of health, limb or life caused by any incident, accident, and natural, technological or deliberate accident or long-term exposure to harmful emissions or radiation, with the decrease of the ALE in good health of individuals affected by this event.

The basis of the model is the basic (unperturbed) table of mortality/survival of the population living in the immediate vicinity (destruction zone) of the critical infrastructure or the potentially dangerous object (PDO). The population of a district, town, region or country is taken as the original cohort that allows evaluating the relative contribution of the accident to the decrease of life expectancy.

The main specific demographic feature of an accident/ catastrophe/ terrorist attack is that they occur at random times and lead to death, injury, disability and hospitalization of an unknown number of people of different sex, age, health and qualification. These events are random disturbances of the mortality/survival table, which unambiguously lead to an ALE decrease, for any size of the initial cohort. This allows comparing two mortality tables, both compiled for the same population: one before and the other after the accident and thereby estimate the magnitude of the ALE reduction.

The modern theory of demographic statistics is used to determine how many person-years are irrevocably lost due to the premature death of the people involved in the accident.

The paper considers three cases: the impact of one person's death on ALE of the cohort in consideration; the effect of the death of several people of the same age on ALE of the cohort in consideration; and the impact of

the death of several people of different ages on ALE of the cohort in consideration.

## **Alternative Measures of Average Life Span, by Socio-Economic Position in England and Wales: a 30 Years Comparison**

**Asim Butt, Chris White**

*Office for National Statistics (ONS), UK*

Life expectancy at birth (mean age at death) is well understood and widely used to inform population ageing, relevant to health and social care provision, extending working lives and setting pensions policies. However, because of improvement in mortality at younger ages, usage of the mean age at death has limitations. The median and modal ages at death have emerged as alternative measures of longevity. This research used all three measures to compare the size of the socioeconomic inequality over a thirty year period using NS-SEC classes. Data from the ONS longitudinal study was applied to a Chiang abridged life table closed at 95+. The median and mode were approximated using the formula for grouped data, based on expected deaths available from the life table.

Across all three measures, age at death improved markedly across all NS-SEC classes for men and women alike; however, the mode showed the strongest improvement for men in Routine occupations. For instance, the Routine class modal age at death for men improved by 9.2 years (75.2 to 84.4 years) between 1982-86 and 2007-11; the median age at death increased by 5.83 years (73.5 to 79.3 years) and mean age at death by 5.80 years (70.8 to 76.6 years). This shows most routine men died between ages 84 and 85, and half the cohort was still alive at age 79.3, somewhat higher than their life expectancy. As deaths are concentrating at older ages the modal and median age at death adds value to the traditional longevity measure.

**Keywords:** Most common age at death, modal age at death, Social class

## The Limit of Human Performance in Triple Jump: an approach with Extreme Value Modelling

Frederico Caeiro<sup>1</sup>, Manuela Oliveira<sup>2</sup>, and Domingos Silva<sup>3</sup>

<sup>1</sup> Universidade Nova de Lisboa, Faculdade de Ciências e Tecnologia & Centro de Matemática e Aplicações (CMA), 2829-516 Caparica, Portugal

<sup>2</sup> Universidade de Évora, CIMA, Portugal

<sup>3</sup> Universidade de Évora, CIMA, Portugal

Emil Gumbel was the first to use extreme value models in statistics applications. He considered a sample of block maxima obtained from non-overlapping blocks (the block length is usually chosen as one year) of a series of independent and identically distributed random variables. Although the block maxima method has proved to be useful in diversified situations, it has also been criticized since we are wasting information by using only the observed maxima from each block. To use more information about the tail of the model underlying the data, the block maxima method was more recently extended to the  $r$ -largest order statistics method. The choice of the number  $r$  of largest order statistics from each block must be made with careful, due to the usual bias and variance trade-off.

In this work we use the  $r$ -largest order statistics method to study the limit of human performance in the Triple Jump event. Our results indicate a negative extreme value index and thus a finite right endpoint for the extreme value model.

**Keywords:** Extreme value theory, Block maxima, Generalized extreme value distribution.

**Acknowledgements:** *research partially funded by FCT - Fundação para a Ciência e a Tecnologia, Portugal, through the projects UID/MAT/04674/2013 (CIMA: Centro de Investigação em Matemática e Aplicações) and project UID/MAT/00297/2013 (Centro de Matemática e Aplicações).*

## Calibration of Multiscale Two-Factor Stochastic Volatility Models: A Second-Order Asymptotic Expansion Approach

Betuel Canhanga<sup>1</sup>, Anatoliy Malyarenko<sup>2</sup>, Ying Ni<sup>2</sup>,  
Milica Rancic<sup>2</sup>, Sergei Silvestrov<sup>2</sup>

<sup>1</sup>Faculty of Sciences, Dept of Mathematics and Computer Sciences, Eduardo Mondlane University, Mozambique, <sup>2</sup>Division of Applied Mathematics, School of Education, Culture and Communication, Mälardalen University, Sweden

The development of financial markets imposes more complex models on the option pricing problems. On the previous papers by the authors, we

consider a model under which the underlying asset is driven by two independent Heston-type stochastic volatility processes of multiscale (fast and slow) mean-reverting rates and we compute an approximate solution for the option pricing problem, using asymptotic expansion method.

In the present paper, we aim to calibrate the model using the market prices of options on Euro Stoxx 50 index and an equity stock in the European market. Our approach is to use the market implied volatility surface for calibrating directly a set of new parameters required in our second-order asymptotic expansion pricing formula for European options. This second-order asymptotic expansion formula provides a better approximation formula for European option prices than the first-order formula, as explained in an earlier work of the authors.

**Keywords:** Option pricing model, asymptotic expansion of option price, stochastic volatility model, multiscale stochastic volatility, calibration.

## On the Choice of a Model-Based Tree for Ordinal Scores

**Carmela Cappelli, Francesca di Iorio, Rosaria Simone**

*Dept of Political Sciences, Università Federico II, Italy*

Ordinal responses in the form of ratings arise frequently in social sciences, marketing and business applications where preferences, opinions and perceptions play a major role. Main examples concern customers/users' satisfaction analysis where it is common to collect rater's evaluation on a hedonic scale, along with a set of covariates (both categorical and quantitative) that characterize the respondent and/or the item/service. In this framework, the ordinal nature of the response has to be properly taken into account when the interest is in the understanding of different response patterns in terms of subjects' covariates. In this spirit, a model-based tree procedure for ordinal scores is illustrated: its structure is based on a class of mixture models for ordinal rating data that implies a twofold analysis in terms of feeling and uncertainty and effective graphical visualization of results. The flexibility of the chosen modelling framework entails that the splitting criterion can be customized according to the purposes of the study and the available data, without disregarding uncertainty. Thus, the selection of variables yielding to the best partitioning results is driven by fitting measures or classical likelihood and deviance measurements, for instance. The contribution proposes to investigate the features of varying decision rules and thus implicitly addresses the problem of selecting the model-based tree that provides the most adequate and satisfying overview of response profiles. Comparison with alternative model-based approaches is also outlined.

**Keywords:** Ordinal Data, Model-based Trees, Splitting Criteria.

**References:**



- [1]. C. Cappelli, R. Simone and F. di Iorio. Model-based trees to classify perception and uncertainty: analyzing trust in European institutions, under review, 2017.
- [2]. A. D'Elia, A. and D. Piccolo. A mixture model for preference data analysis. *Computational Statistics & Data Analysis*, 49, 917-934, 2005.
- [3]. G. Galimberti, G. Soritti, and M. Di Maso, Classification Trees for Ordinal Responses in R: The RpartScore Package, *Journal of Statistical Software*, 47, 1-25, 2012.
- [4]. R. Picarretta. Classification trees for ordinal variables, *Computational Statistics*, 23, 407-427, 2008.

## Competing Markov Manpower Models and states Based on Seniority

**Philippe Carette<sup>1</sup>, Marie-Anne Guerry<sup>2</sup>**

<sup>1</sup>*Dept of General Economics, Ghent University, Belgium,*

<sup>2</sup>*Dept Business Technology and Operations, Vrije Universiteit Belgium*

Markov chains are useful in modeling manpower systems. The states of a Markov manpower model are then personnel groups that are homogeneous regarding the transition probabilities. In case transition probabilities depend on seniority, alternative approaches can be considered in striving to improve the model. One approach is to implement a more detailed specification of the states based on a length of stay criterion. In doing so, two competing Markov models can be taken into consideration depending on whether or not, the states are additionally specified by seniority. In aiming to select an adequate manpower model, the present paper compares the performance of the two resulting Markov models for which the set of states are selected differently. The goodness of fit is measured based on likelihood-computation. Properties of the likelihood function are proved to find insights regarding the effect of incorporating a time duration variable in defining the states of the Markov manpower model.

**Keywords:** Markov chain, manpower model, duration effect, state selection, likelihood function, model performance.

### References:

- [1]. D.J. Bartholomew, A.F. Forbes, S.I. McClean. *Statistical techniques for manpower planning*, second ed. John Wiley & Sons, Chichester, 1991.
- [2]. H.H.Z. Thom, C.H. Jackson, D. Commenges, and L.D. Sharples. State selection in Markov models for panel data with application to psoriatic arthritis. *Statistics in medicine*, Wiley Online Library, 34, 16, 2456-2475, 2015.

## Modeling of Mortality in Elderly by Prostate Diseases in Brazil

João Batista Carvalho<sup>1</sup>, Neir Antunes Paes<sup>2</sup>

<sup>1</sup>Dept of Statistics, Federal University of Campina Grande, Brazil,  
Federal University of Paraíba, Brazil, <sup>2</sup>Health and Decision Modeling  
Postgraduate Course, Federal University of Paraíba, Brazil

Northeastern Brazil can be considered as one of the less developed regions of the country. With a population of 56 million inhabitants in 2016, it features deficient vital statistics, which hinders the ability of management and development of public health policies. Aiming at identifying explanatory factors of life conditions and vulnerability of one of the major causes of mortality in elderly, the prostate diseases, a cross-sectional ecological study to the year 2010, using micro-data information, was adopted to the 188 micro regions of the northeastern of Brazil. The following steps were performed: 1) The death data were corrected by underreporting, and ill defined causes as well garbage codes were redistributed to the corresponding prostate deaths by applying appropriate techniques; 2) The evolution of corrected mortality rates from 2010 to 2015 was performed; 3) The Structural Equations Modeling (SEM) was applied. The outcome mortality rates due to prostate disease were observed directly, and a set of life conditions and vulnerability indicators regarding to health, education, income and environmental conditions were used indirectly as latent variable. From the Health Ministry were extracted mortality data. This study traced a regional overview of the mortality due to prostate diseases among old men in the northeast of Brazil, pointing out distinct realities by using the micro-data of the late Census 2010. The SEM proved to be highly sensitive with significance in the measurement model for some relevant latent variables that can subsidize political planning. In view of the high rates observed and an aging process still rising in the Northeast, is expected that these levels should increase even more in the near future, aggravating an already worrying picture.

**Keywords:** prostate diseases, mortality in the elderly, mortality by causes

## **Determining the Structure and Assessing the Psychometric Properties of Multidimensional Scales Constructed from Ordinal and Pseudo-Interval Items**

**Anastasia Charalampi<sup>1</sup>, Catherine Michalopoulou<sup>1</sup>,  
Clive Richardson<sup>2</sup>**

*<sup>1</sup>Dept of Social Policy, Panteion University of Social and Political Sciences, Greece, <sup>2</sup>Dept of Economic and Regional Development, Panteion University of Social and Political Sciences, Greece*

Determining the structure and assessing the psychometric properties of multidimensional scales before their application is a prerequisite of scaling theory. This involves splitting randomly a sample of adequate size into two halves and first performing exploratory factor analysis (EFA) on one half-sample in order to assess the construct validity of the scale. Secondly, this structure is validated by carrying out confirmatory factor analysis (CFA) on the second half. As in any statistical analysis -whether univariate, bivariate or multivariate - the first and most important consideration is to ascertain the level of measurement of the input variables, in this instance the defining items of the scale. This guides the correct choice of the methods to be used. In this paper, we carry out the investigation and assessment of the 2006 European Social Survey (ESS) six-dimensional instrument of wellbeing for Germany and Netherlands when items are considered as ordinal and pseudo-interval.

**Keywords:** Exploratory factor analysis, Confirmatory factor analysis, Reliability, Construct validity, European Social Survey

## **Using Stochastic Health State Function to Forecast Healthcare Demand and Healthcare Financing: Evidence from Singapore**

**Ngee Choon Chia and Shu Peng Loh**  
*Dept of Economics, National University of Singapore*

Typically, healthcare financing for an ageing population requires projections on healthcare demand and cost. However, projecting healthcare demand based on projected elderly does not consider changes in population health state over time. This paper proposes a new approach to forecast health variables using a stochastic health state function and the well-established Lee-Carter stochastic mortality model. With the estimated health state at each age over time, we project the hospitalization rate, healthcare demand and financing cost for Singapore using historical life tables and hospital admission data. Our findings show that while hospital insurance claims increase due to an aging population,

improving health state could save costs from hospital insurance claims. This has policy implications: more attention should be given to preventive healthcare such as health screening to improve the overall health state of the population.

**Keywords:** Ageing, Longevity, Health State, Hospitalisation Demand, Healthcare Financing

## **Opening Up the Statistical Genetics Toolbox to the Longitudinal Study of Reproductive Ageing**

**Christodoulou, M., Brettschneider, J., Steinsaltz, D.**  
*Dept of Statistics, University of Oxford, UK*

The correlation between female fertility patterns and ageing is both complex and poorly understood. Menopause timings and the underlying mechanisms have long fascinated biologists, demographers, and medical researchers; however, the available statistical tools have constrained the questions that were feasible to address. Although the timings of both menarche (timing of first menstruation) and menopause have been studied in terms of heritability and environmental influences, the two events have rarely been considered simultaneously in a genetic context. Furthermore, the reproductive pathways of the studied population have not been taken into account as factors affecting the fertility timings. Our work aims to answer questions such as: How does the timing of pregnancies affect when menopause occurs? Does the reproductive window (the timing between menarche and menopause) respond to an individual's reproductive events? How can we best describe population-wide variations in patterns of senescence, and their genetic basis? There is a wealth of unexploited information in the British Cohort Studies that could shed light on these questions, if two prerequisites were accomplished: first, the extraction of complete fertility profiles from the often chaotic longitudinal data for cohort members; and second, the development of analytical tools that can extend joint modelling approaches to include genetic information, and can incorporate complex lifetime event processes. We describe analytical tools that can be used for phenomena with complex longitudinal, to extract clearer signals of individual ageing and better estimate heritability. The methods are being applied to the 1958 British Birth Cohort as a proof of concept.

**Keywords:** fertility, menopause, longitudinal studies, joint modelling

## Gender Differences and Control for Longevity Risk with indexed Pension Systems

Mariarosaria Coppola<sup>1</sup>, Maria Russolillo<sup>2</sup>, Rosaria Simone<sup>1</sup>

<sup>1</sup>*Dept of Political Sciences University of Naples Federico II, Italy,*

<sup>2</sup>*Dept of Economics and Statistics, University of Salerno, Italy*

The goal of the contribution is to assess the different effects that the assumption of an indexation mechanism for retirement age could imply on gender policies in order to deal with longevity risk. Indeed, the ageing of population is becoming an issue of foremost pressure, especially in countries like Italy experiencing low fertility rates. In this regard, it is well known that longevity risk assumes different strengths on male and female populations, with decrease in mortality rates evolving at different speed over time. In order to control the longevity risk and foster the understanding of such differences, the contribution considers an indexation mechanism based on the expected residual life to adjust the retirement age and keep a constant Expected Pension Period Duration (EPPD). A comparative analysis is then pursued for the Italian male and female populations: forecasts of competing stochastic mortality models are then subject to model assembling techniques in order to balance fitting performances and uncertainty related to model selection.

**Keywords:** Longevity Risk, Model Assembling, Gender Differences, Indexed Retirement Scheme.

### References:

- [1]. A.G. Benchimol et al. Model uncertainty approach in mortality projection with model assembling methodologies, Working Papers. Statistics and Econometrics. WS Universidad Carlos III de Madrid Departamento de Estadística DES, 2016.
- [2]. M. Coppola, M. Russolillo and R. Simone. The Impact of Mortality Projection Models in Case of Flexible Retirement Schemes. Demography and Health Issues - Population Aging, Mortality and Data Analysis - The Springer Series on Demographic Methods and Population Analysis, pp. 347{358, 2018.
- [3]. E. Bisetti and C.A. Favero, 2014: Measuring the impact of longevity risk on pension Systems: the case of Italy. North American Actuarial Journal 18(1), 2014.
- [4]. M. Denuit M., S. Haberman S. and A.E. Renshaw. Longevity-contingent deferred life annuities. Journal of Pension Economics and Finance 14 (3): 315-327, 2015.
- [5]. A.R. Villegas, V. Kaishev, P. Millossovich. StMoMo: An R Package for
- [6]. Stochastic Mortality Modelling. Available at <https://cran.r-project.org/web/packages/StMoMo/vignettes/StMoMoVignette.pdf> 2016.

## Content of Memory, Dream and Imagination: Exploring Processes in a Multilevel Approach

Franca Crippa, Angela Tagini

*Dept of Psychology, Italy*

Multilevel models have found application in several areas of behavioural studies, when data are clustered, particularly in the case of more observations for each person, i.e. when repeated measures are clustered within subjects. In autobiographical memory, subjects provide specific details and characteristics for each past event and for each imagination of a future scenario, they have resumed with relation to a common stimulus. Traditionally, most studies have been exploring the relationship between past and future by aggregating data across all events generated by an individual. In this way, the investigation of links between past memories and future imagination disregards the correlational structure within the events. Fairly recently, though, yet, the analytical advantage of the multilevel approach has been shown in mental time travel from past to future, in designing new studies or even reanalysing datasets from previous ones. This has allowed researchers to explore the relationship between fundamental aspects of the process under investigation, for instance between semantic and episodic details within events. We aim to adapt the aforementioned multilevel modelling framework to the investigation of relations underlying past collection and future imagination with dream, within an ongoing research where data have been collected in line with a coherent theoretical approach.

**Keywords:** autobiographical memory, multilevel model, repeated measure, fixed effect, random effect.

### References:

[1] Devitt, A. L., Addis, D. R., & Schacter, D. L. (2017). Episodic and semantic content of memory and imagination: A multilevel analysis. *Memory & cognition*, 45(7), 1078-1094.

## A Multi-State Model for Studying the Demography of Species Providing Extended Parental Care: a Case Study on Polar Bears

**Sarah Cubaynes<sup>1,2</sup>, Jon Aars<sup>3</sup>, Nigel Yoccoz<sup>4</sup>, Roger Pradel<sup>1</sup> and Olivier Gimenez<sup>1</sup>**

<sup>1</sup>MMDN, INSERM U1198, EPHE/PSL & University of Montpellier, France

<sup>2</sup>Centre d'Ecologie Fonctionnelle et Evolutive, UMR 5175, Montpellier, France

<sup>3</sup>Norwegian Polar Institute, Division for Ecology, Tromsø, Norway

<sup>4</sup>UiT The Arctic University of Norway, Dept of Arctic and Marine Biology, Tromsø, Norway

In species providing extended parental care, offspring are dependent on both or one parent for a prolonged period of time during which parent(s) (often the mother) will miss out breeding opportunities. Such reproductive strategy is common in long-lived mammal species in which offspring' survival beyond independence rely on learning essential life skills from parent(s) (e.g. hunting, orientation, or nest building).

In those species, parental investment in reproduction is high and prolonged in time, parent(s) breeding attempt depends on past and present offspring number, age and condition, while offspring survival is also conditioned upon parent(s) survival and condition over several years. As a consequence, parent(s) breeding frequency and inter-birth interval can be highly variable and life histories are interdependent within family units until offspring independence.

Existing demographic models are often too simplistic to model the demography of species providing extended parental care, because they do not explicitly include these dependencies. In addition, it is often impossible to complete an exhaustive monitoring of individuals from birth to death in wild animal populations and we have to deal with the problem of incomplete or missing data.

To overcome these challenges, we propose a capture-recapture multi-state model specifically designed to estimate demographic parameters (parents and offspring survival, inter-birth interval, breeding frequency, litter size) in species providing extended parental care. We apply our model to study female Polar bears reproductive strategies using a long-term monitoring database collected from the Svalbard population. We also evaluate the model performance on simulated data. We discuss potential questions that could be addressed with this model, and possibilities for further developments of the model.

## **A Non-Homogeneous Semi-Markov Model for the Calculation of the claim reserving of a Health Fund**

**Guglielmo D'Amico<sup>1</sup>, Fulvio Gismondi<sup>2</sup>, Montserrat Guillen<sup>3</sup>, Jacques Janssen<sup>4</sup>, Raimondo Manca<sup>5</sup>, Ernesto Volpe di Prignano<sup>5</sup>**

<sup>1</sup>Dept of Pharmacy, University "G. d'Annunzio" of Chieti-Pescara, Italy,

<sup>2</sup>University "Guglielmo Marconi" of Rome, Italy, <sup>3</sup>Departament d'Econometria, Estadística i Economia Aplicada University of Barcelona, Spain,

<sup>4</sup>Université Libre de Bruxelles, Belgium,

<sup>5</sup>MEMOTEF Dept, University of Rome "La Sapienza", Italy

This paper will present a model that gives the possibility of the calculation of the expenses of a health insurance fund without considering the administrative costs. To get this result we subdivide the insured people in cohorts. We will follow, for each cohort, the evolution of the costs of claims of the fund.

We will work in Non-Homogenous Semi-Markov setting considering also the ReWard Process with Backward recurrent time (see Stenberg et al. (2007) and (D'Amico et al. (2010))).

The time variable that we consider is the age. Indeed, it is easily understood that an illness has different evolutions depending on the age of the insured. It is natural supposing that the illness evolution is function of the different age and illness and it will not change too much in different calendar times at a given age.

## **A Probabilistic Model of wind Farm Power Generation via Copulas and indexed semi-Markov models**

**Guglielmo D'Amico<sup>1</sup>, Giovanni Masala<sup>2</sup>, Filippo Petroni<sup>2</sup>**

<sup>1</sup>Dept of Pharmacy, University "G. d'Annunzio" of Chieti-Pescara, Italy,

<sup>2</sup>Dept of Business and Economics, University of Cagliari, Italy

In this paper we consider the problem of modelling the wind power production of a wind farm composed of a given number of wind turbines. As it is well known, the comprehension of stochastic properties of the total produced energy cannot be obtained simply considering the produced power of a single turbine with the total number of turbines. The reasons are mainly due to the so-called shear effect and to the morphological nature of the territory where the wind farm has been installed. For these reasons, it is interesting and crucial the development of a complete model that is able to correctly reproduce and forecast the power production of the whole wind farm. To this end we describe the stochastic production of energy of each turbine using an indexed semi-Markov chains (ISMC). This choice is motivated by recent research articles where the authors have



demonstrated that ISMC model is able to reproduce the statistical properties of power production of a single wind turbine, see [2-7].

The ISMC model provides a very general approach that encompasses both semi-Markov processes and Markov chain based models. The superiority of the ISMC model resides in its accurate probabilistic description of the wind power evolution which accounts for the serial dependence of the wind power time series by incorporating past events (times and sizes of past power production) through an index process that increases the memory of the process.

The modelling of the whole wind farm is executed by introducing a dependence structure among the considered ISMC that describe the wind power of each single wind turbine. The dependence structure is described by introducing copula functions. The advanced model is a modification of that presented in [1] where weighted-indexed semi-Markov chain model with copulas was applied to the description of financial performance of a portfolio of financial assets varying on a high frequency scale.

A real application of the proposed multivariate model is performed on real data of energy produced have been analysed and compare with synthetic data obtained by implementing Monte Carlo simulation and the multivariate ISMC model.

**Keywords:** wind energy, indexed-semi-Markov chain, copula.

## References

- [1]. G. D'Amico and F. Petroni. "Multivariate high-frequency financial data via semi-Markov processes" submitted to European Journal of Operational Research
- [2]. G. D'Amico, F. Petroni and F. Prattico. "Wind speed modeled as a semi-Markov process with memory" *Environmetrics* 24(6), 2189-2201 (2013).
- [3]. G. D'Amico, F. Petroni and F. Prattico. "Wind speed and energy forecasting at different time scales: A nonparametric approach" *Physica A: Statistical Mechanics and its Applications* 406, 59-66 (2014).
- [4]. G. D'Amico, F. Petroni and F. Prattico. "Wind speed prediction for wind farm applications by Extreme Value Theory and Copulas" *Journal of Wind Engineering and Industrial Aerodynamics* 145, 229-236 (2015).
- [5]. G. D'Amico, F. Petroni and F. Prattico. "Reliability measures for indexed semi-Markov chains applied to wind energy production" *Reliability Engineering and System Safety* 144, 170-177 (2015).
- [6]. G. D'Amico, F. Petroni and F. Prattico. "Economic performance indicators of wind energy based on wind speed stochastic modeling" *Applied Energy* 154, 290-297 (2015).
- [7]. G. D'Amico, F. Petroni and F. Prattico. "Insuring wind energy production" *Physica A: statistical mechanics and its applications* 467, 542-553 (2017).

## Optimal Provision of a Dispatchable Energy Source for Wind Energy Management: Dependence on the Wind Energy Model

Guglielmo D'Amico<sup>1</sup>, Filippo Petroni<sup>2</sup>, Fulvio Gismondi<sup>3</sup>

<sup>1</sup>Dept of Pharmacy, University "G. d'Annunzio" of Chieti-Pescara, Italy,

<sup>2</sup>Dept of Business and Economics, University of Cagliari, Italy,

<sup>3</sup>Dept of Economic and Business Science, "Guglielmo Marconi" University, Italy

Wind energy is assuming even more importance in the production of electricity. The share of production due to wind is continuously increasing in time although there are still relevant problems that affect this industry. The most important limitation for a further development of the wind energy industry concerns the variability of the wind speed phenomenon.

The problem of the wind speed volatility has been approached mainly by energy storage systems; that is, by storing a surplus of energy to be used for compensating an eventual future deficit of production. More recently an insurance contract between the wind energy producer (WEP) and a dispatchable energy producer (DEP) has been proposed as a mean to manage the uncertainty of the wind speed.

In this paper we assume that the WEP is also able to produce energy by means of gas and that he has agreed to furnish a given quantity of energy  $K$ . An insufficient production of energy determines a cost to be suffered because penalties apply. However, an excess of production is lost. Therefore, the energy producer should determine the optimal quantity of energy to be produced with gas that added to the uncertain wind energy production maximize his expected profit. The problem is solved under different hypothesis on the wind energy model. First, the wind energy production is modelled by a simple sequence of i.i.d. random variables, then a Markov chain model is used and finally semi-Markov based models of wind energy are applied. The results show the dependence of the optimal policy on the different models of wind energy and therefore highlight the importance of using an appropriate model of wind energy. The application is performed on real data of energy produced by a wind turbine E-48 ENERCON of rated power 800kW.

**Keywords:** wind energy, optimal provision, semi-Markov.

## Reducing Wind Power Uncertainty Using Insurance Contracts

**Guglielmo D'Amico<sup>1</sup>, Filippo Petroni<sup>2</sup>, Fulvio Gismondi<sup>3</sup>**

*<sup>1</sup>Dept of Pharmacy, University "G. d'Annunzio" of Chieti-Pescara, Italy, <sup>2</sup>Dept of Business and Economics, University of Cagliari, Italy, <sup>3</sup>Dept of Economic and Business Science, "Guglielmo Marconi" University, Italy*

The share of electricity production due to wind is continuously increasing although relevant problems still affect this industry. The most important limitation for a further development of the wind energy industry concerns the variability of the wind speed phenomenon. The problem of the wind speed volatility has been approached mainly by energy storage systems; that is, by storing a surplus of energy to be used for compensating an eventual future deficit of production. More recently an insurance contract between the wind energy producer and a dispatchable energy producer has been proposed as a mean to manage the uncertainty of the wind speed. In this paper we extend previous results involving the use of insurance contracts by considering the dependence existing between electricity prices and wind energy production. The dependence structure is modeled using an appropriate copula function and we show the impact of this dependence on the fair premium that the wind power supplier has to pay in order to hedge the risk of inadequate output of electricity at any time. Recursive type equations are obtained for the prospective mathematical reserves of the insurance contract and for their higher order moments. The model and the validity of the results are illustrated through a numerical example.

**Keywords:** wind power production, indexed semi-Markov chains, copula function.

## Entropy Analytics of GDP/Capita Distribution in E.U.-28 and Eurozone-19 Members States

**Yiannis Dimotikalis**

*Dept of Business Administration, Technological Education Institute of Crete,  
Agios Nikolaos, Crete, Greece*

European Union is an association of 28 participant countries for the common economy and social improvement of their societies. The distribution of GDP per capita is a main index of common interest on economic convergence. In this study analyzed the GDP distribution in E.U.-28 and Eurozone-19 by entropic approach in the period 1995-2016 where the data are available. The inequality of income distribution is the main topic and convergence is examined using Theil index and Kullback-Leibler divergence of Population, GDP and GDP per Capita data. A

Theoretical analysis is performed about the implications of the entropy approach and the possible final equilibrium states appears to predict and resulted. In some way the desirable uniform distribution of perfect equilibrium is an ideal case of the general exponential or binomial distributions resulting from entropy maximization under certain constraints. The whole E.U.-28 economy compared with U.S.A., Russia, Japan and China GDP/Capita of the period 1995-2016, using Theil inequality Index. Some scenarios for the future of convergence/divergence of those Big 5 economies presented and examined for the next decade, based on estimations for economic development of OECD and Word Bank. Some theoretical problem introduced and discussed about the entropy method and limitations of this approach.

**Keywords:** Maximum Entropy, Geometric Distribution, Binomial Distribution, Nonlinear Constrained Optimization, Theil Inequality Index, Kullback-Leibler Divergence, Income Inequality.

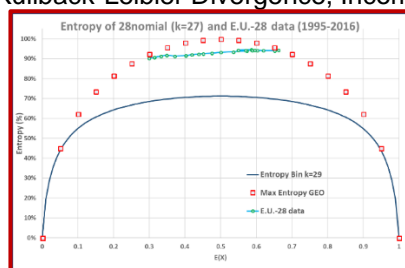


Fig. 1: Entropy of E.U.-28 (1995-2016)

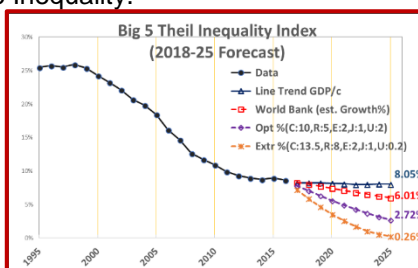


Fig. 2: Big 5 Inequality Index Forecast Scenarios

## The Necessity to Use Advanced Methods of Analysis in Order to Approach Complex Phenomena: The Example of the Territorial Attractiveness of the Municipalities of Greece

Marie-Noelle Doquenne<sup>1</sup>, Stamatina Kaklamani<sup>2</sup>

<sup>1</sup>Dept of Planning and Regional Development, University of Thessaly, Greece,

<sup>2</sup>Dept of Philosophy and Social Studies, University of Crete, Greece

The aim of the proposed work is to formulate and evaluate advanced analysis methods to assess the degree of attractiveness of the territorial entities of a region or even a country. Understanding the multiple factors that, in synergy, enable spatial attractiveness was based on the precise identification of the main dimensions of attractiveness and their components. Consequently, each component is studied by defining a set of variables. The "measurement" of the variables is done through the definition of appropriate and understandable indicators adapted to the

surveyed spatial scale but under the limitation of availability-accessibility to the necessary data. Factor Analysis helps to highlight the interactions between the various measures proposed through the extraction of composite indicators (principal components). Consequently, through the principal components and classification methods, the various spatial motifs of attractiveness are detected. The proposed method and its evaluation are controlled through its implementation in all municipal units (formerly Kapodistrian municipalities) of the countryside of Greece, that is, approximately 900 territorial entities.

**Keywords:** Greece countryside, attractiveness, Factor Analysis, Classification Methods

## Estimating Functions in The Presence of a Nuisance Parameter-The Multivariate Case

**Douli Marigo**

*Dept of Statistics and Insurance Sciences,  
University of Piraeus, Greece*

In this work lower bounds for the estimation of a real-valued parameter  $\alpha$  of parametric models in the presence of a nuisance parameter  $\beta$  are presented. In particular, an optimal estimating function which provides an estimator for with the smallest possible variance was developed, while for the elimination of the nuisance parameter an estimator which is independent of the sample was used. The resulting estimating function is better as compared with Stein's adaptive score function.

**Keywords:** Estimating functions; nuisance parameters; adaptive score function.

### References:

- [1]. Bhapkar, V. P. Sufficiency, ancillarity and information in estimating functions. Estimating Functions. Edited by Godambe, V. P. Oxford Science Publications, Oxford. 241-254, 1991.
- [2]. Bickel, Klaassen, Ritov and Welner. Efficient and adaptive estimation for semiparametric models. John Hopkins University Press, Baltimore, 1993
- [3]. Zaigraev A., Podraza- Karakulska A. Maximum integrated likelihood estimator of the interest parameter when the nuisance parameter is location or scale. Statistics and Probability Letters 88, 99-106, 2014
- [4]. Godambe, V. P. An optimum property of the regular maximum likelihood estimator. Ann. Math. Statist. 2 1208-12, 1960.
- [5]. Liang, K. On information and ancillarity in the presence of a nuisance parameter. Biometrika 70. 607-612, 1983.

## Monitoring the Number and the Spatial Distribution of Disease Events Using Convex Hulls and Control Charting

Polychronis Economou<sup>1</sup>, Sotiris Bersimis<sup>2</sup>, Athanasios Sachlas<sup>2</sup>

<sup>1</sup>Dept Civil Engineering, University of Patras, Greece,

<sup>2</sup>Dept of Statistics & Insurance Science, University of Piraeus, Greece,

To timely and efficiently detect disease outbreaks one should consider both spatial and temporal dimensions. Of interest are global changes in the number of new disease events on time and/or hotspots of disease events which may evolve into outbreaks. A key assumption in biosurveillance is that under normal conditions events are uniformly distributed in the plane. In this work, a new two-step monitoring procedure is proposed with which the number of disease events through control charting and the spatial distribution of disease events via convex hulls are monitored. The numerical illustration showed an remarkable performance of the new test under different outbreaks scenarios.

**Keywords:** Biosurveillance, Control Charts, Convex Hulls, Monitoring

## Statistical Analysis of Type-II Progressive Hybrid Censoring for Generalized Half Normal Distribution

Farha Sultana, Yogesh Mani Tripathi

Dept of Mathematics, Indian Institute of Technology Patna, India

Censoring occurs quite naturally in reliability and life testing experiments. Estimation problems based on censored data have found wide applications reliability studies. In literature different censoring methodologies have been proposed. Kundu and Joarder [1] introduced Type-II progressive hybrid censoring scheme and studied a one-parameter exponential distribution under this scheme. Authors used classical and Bayesian approaches to estimate the unknown parameters. Both point and interval estimates are discussed. One may also refer to Childs, Chandrasekar and Balakrishnan [2] for some further results on this censoring.

In this paper we consider estimation of unknown parameters of a Generalized Half Normal distribution under Type-II progressive hybrid censoring scheme. We obtain maximum likelihood estimates of parameters using an expectation-maximization algorithm. Asymptotic intervals are constructed from the Fisher information matrix. We also obtain Bayes estimates under the squared error loss function using the Lindley method and importance sampling procedure. The problem of prediction is also considered under Bayesian framework. We obtain

prediction estimates of censored observations and construct associated prediction intervals. The performance of proposed methods is compared numerically using simulations and comments are obtained based on this numerical study.

Finally we analyze a real data set for illustrative purposes.

### References

- [1]. D. Kundu and A. Joarder. Analysis of Type-II progressively hybrid censored data. *Computational Statistics and Data Analysis*, 50, 2509-2528, 2006.
- [2]. A. Childs, B. Chandrasekar and N. Balakrishnan, Exact likelihood inference for an exponential parameter under progressive hybrid censoring schemes. *Statistical models and methods for biomedical and technical systems*. Birkhauser, 323-334, 2004.

## Coefficient of Variation in Life tables

**Nikolaos Farmakis, Dimitrios Seirinopoulos**

*Aristotle University of Thessaloniki, Greece*

The paper will present the basic concepts of gender and age composition of a population, population pyramid and fertility, mortality and migration indicators that study a population and analyze the survival tables and their graphs with the use of the coefficient of variation.

**Keywords:** Survival tables, mortality, population pyramid, coefficient of variability

## General Bahr-Esseen Inequalities and their Applications

**István Fazekas, Sándor Pecsora**

*University of Debrecen, Faculty of Informatics, Hungary*

For independent random variables with finite  $p$ th moment and mean zero the  $p$ th von Bahr-Esseen moment inequality was obtained by von Bahr and Esseen[3]. The von Bahr-Esseen moment inequality was proved for pairwise independent random variables in Chen *et al.*[2].

In this paper we shall show that the Bahr-Esseen inequality holds with exponent  $p$ , if it holds with exponent  $q$  for the truncated and centered random variables if  $q > p$ . To obtain the result we shall apply the method presented in Chen *et al.*[2].

The notion of acceptability was introduced in Antonini *et al.*[1]. In this paper we shall prove that if the truncated random variables are acceptable, then an exponential inequality is true, which implies Rosenthal's inequality and also von Bahr-Esseen's moment inequality.

Then we apply the results to obtain weak and strong laws of large numbers and complete convergence for random variables without assuming independence.

**Keywords:** Moment inequality, law of large numbers.

#### References

- [1]. R. G. Antonini, Y. Kozachenko and A. Volodin: Convergence of series of dependent  $\phi$  sub-Gaussian random variables. J. Math. Anal. Appl. 338 (2008), no. 2, 1188-1203.
- [2]. Pingyan Chen, Peng Bai and Soo Hak Sung: The von Bahr-Esseen moment inequality for pairwise independent random variables and applications. J. Math. Anal. Appl. 419 (2014), no.2, 1290-1302.
- [3]. B. von Bahr and C. G. Esseen: Inequalities for the  $r$ th absolute moment of a sum of random variables,  $1 \leq r \leq 2$ . Ann. Math. Statist. 36 (1965), no.1, 299-303.
- [4]. I. Fazekas and S. Pecsora : General Bahr-Esseen inequalities and their applications. Journal of inequalities and applications (2017) 191.

## Some Properties of the Multivariate Generalized Hyperbolic Laws

**Stergios B. Fotopoulos<sup>1</sup>, Venkata K. Jandhyala<sup>2</sup>, Alex Paparas<sup>1</sup>**

<sup>1</sup>*Dept of Finance and Management Science, Washington State University, Pullman, WA 99164-4746, USA*

<sup>2</sup>*Dept of Mathematics & Statistics, Washington State University, Pullman, WA 99164-4113, USA*

The purpose of this study is to characterize multivariate generalized hyperbolic (MGH) distributions and their conditionals by considering the MGH as a subclass of the mean-variance mixing of the multivariate normal law. The essential contribution here lies in expressing multivariate generalized hyperbolic densities by utilizing various integral representations of the Bessel function. Moreover, in a more convenient form these modified density representations are more advantageous for deriving limiting results. The forms are also convenient for studying the transient as well as tail behavior of multivariate generalized hyperbolic distributions. The results include the normal distribution as a limiting form for the MGH distribution. This means the MGH model can be considered for modeling not only high frequency data but also for modeling low frequency data. This is against the currently prevailing notion that the MGH model is relevant for modeling only high frequency data.

**Keywords:** Scale mixture of multivariate distributions; generalized inverse Gaussian distributions; Non Gaussian; Conditional laws.



## On Determining the Value of Online Customer Satisfaction Ratings – a Case-Based Appraisal

**Jim Freeman**

*Alliance Manchester Business School, University of Manchester, UK*

A positive online reputation is one of the most powerful marketing assets a business can possess. Thanks to ubiquitous electronic Word of Mouth (eWOM) the internet has become saturated with online ratings and rankings on consumer satisfaction as a plethora of review sites vie for dominance in the world of reputation marketing.

Unfortunately, the survey design, data collection methods and analytical techniques used for generating such ratings can vary very markedly from review site to review site - not just in transparency but also in value and reliability.

This is especially so in the hospitality industry where hotel review postings in particular often attract debate, suspicion and disagreement. It is against this backdrop that a critical appraisal of the Which? Travel Large UK Hotel Chains 2016 survey operation was undertaken.

**Keywords:** Composite indicators, Factor analysis, Ordinal scores, Voluntary response sampling,

## Credit Portfolio Risk Evaluation with non-Gaussian One-factor Merton Models and its Application to CDO Pricing

**Takuya Fujii, Takayuki Shiohama**

*Tokyo University of Science, Japan*

For measuring credit risks with structural modeling approach, it is often used one-factor Merton models which assume common single risk factor for all counterparties. We extend this model to have non-Gaussian and serially correlated asset returns for both single common risk factor and every obligor's specific process. By using a standard Edgeworth expansion and saddlepoint approximation methods, we derive the approximate default rate distributions. Maximum likelihood estimators are introduced for estimating model parameters, and some credit risk measures are evaluated. Our empirical results illustrate that the proposed non-Gaussian models offer quite different credit risk evaluations compared with those obtained by Vasicek models. Applications for CDO pricing is also discussed.

**Keywords:** Credit risk model, CDO pricing, Edgeworth expansion, Maximum likelihood estimation, saddlepoint approximation

## Aging in the Wild

**Gaillard J.M. & Lemaître J.F.**

*Université Lyon 1; CNRS, France*

Empirical evidence of a decrease in individual performance with increasing age is accumulating in wild populations of vertebrates. Thus, survival, reproductive success and physiological performance all decrease at old ages in the wild, as predicted by evolutionary theories of senescence. However, contrary to the theory, the onset of senescence does often not correspond to the age at first reproduction. Thus, to assess senescence patterns, two metrics, the rate and the onset of senescence, are required. Moreover, the timing and the intensity of senescence is far from constant, but vary strongly both across species in relation to phylogeny, body size, and life style and within species in relation to sex and environmental conditions. The mechanisms involved in such variation are still mostly unknown and are currently under investigation.

## Independence Test for Continuous or Discrete Random Variables

**García, Jesús E., González-López, V.A., Londoño, Sérgio**

*University of Campinas, Brazil*

In this paper we propose a nonparametric test for the hypothesis of independence between random variables. The procedure is derived from the family of tests based on the size of the longest increasing/decreasing subsequence over the observations. See previous members of this family in [1], [2] and [3]. The procedure of this paper does not require the assumption of continuity for the random variables being a natural competitor of the Pearson's Chi-squared test. In the applications we focus on sparse contingency tables, in which the Pearson's Chi-squared test has its validity compromised, see [4].

**Keywords:** Longest increasing subsequence; Test for independence; Discrete statistics.

### References:

- [1].García, Jesús E. and V. A. González-López. "Independence tests for continuous random variables based on the longest increasing subsequence." *Journal of Multivariate Analysis* 127 (2014): 126-146.
- [2].García, Jesús E. et al. "A New Test in the Longest Increasing Subsequence Family of Independence Tests". *Stochastic Modelling and Applications*, 20 (2016): 35-53.
- [3].García, Jesús E. and V. A. González-López. "Independence test for sparse data." *AIP Conference Proceedings*. Vol. 1738. No. 1. AIP Publishing, 2016.
- [4].Agresti, Alan. "Categorical data analysis." John Wiley & Sons, Inc., Hoboken, New Jersey, 2002.

## Similarity between Strains of Zika from Tropical and Subtropical Regions

**García, Jesús E., González-López, V.A.**  
*University of Campinas, Brazil*

The ease with which the Zika virus has developed in tropical and subtropical territories impose on the scientific world the task of deepening the understanding of its mechanisms of growth as well as the task of discover the implications of this virus in the public health. Zika has led many of these countries to borderline situations, since, for instance, the Zika virus is associated with an odd increase of cases of microcephaly and other neurological problems. For example, the high rates at which birth defects occurred in 2015 and 2016 lead to the prediction that Brazil has just a few years to develop specific inclusion policies for these children. Recent studies aim to map the genetic strains of this virus, [1] reported genomic data from several countries, also analyzed the timing and patterns of introductions into distinct geographic regions; the evidence suggests a rapid expansion of the outbreak in Brazil and multiple introductions of outbreak strains into Puerto Rico, Honduras, Colombia, other Caribbean islands, and the continental United States. In the present article, we analyze genomic strains of the Zika virus from patients coming from different tropical and subtropical regions and obtained from the NCBI source. We identify each genomic strain with a finite-order Markov chain, on the base alphabet  $\{a, c, g, t\}$ , we quantify the distance between pairs of strains using a distance  $d$  (see [2]) based on the Bayesian information Criterion, then we use  $d$  to identify the nearest and also the most distant strains. With this methodology we can establish which regions show a similar pattern to the Brazilian strains.

### References:

- [1]. Metsky, Hayden C., et al. "Zika virus evolution and spread in the Americas." *Nature* 546.7658 (2017): 411-415.
- [2]. García, Jesús E. et al. "Linguistic compositions highly volatile in Portuguese." *Cadernos de Estudos Linguísticos* 59.3 (2017): 617-630.

## Examining Credit Scoring Methodologies with Alternative Data

**Panagiota Giannouli, Christos E. Kountzakis**  
*Dept of Mathematics, University of the Aegean, Greece*

The interest in predictive methods and in selection of independent variables for scorecard construction is endless. There has been much advancement in predictive modeling, including the character of independent variables that can be used in credit scoring models and novel

classification algorithms to reliably assess the performance of predictive models. Nevertheless, there are no studies that combine these developments. The objective of this paper is to contribute to both of these areas. To that end, we introduce alternative data to a predictive model which uses only traditional credit behavior data and see if they contribute to model's performance. Subsequently, we conduct a benchmarking study of several novel and traditional classification algorithms in order to compare them according to their performance indicators. This paper provides valuable insights for professionals as they can see novel classification methods and variables in predictive modelling. They can also estimate, based on the results, if it is advantageous to change 'traditional' logistic regression for novel classification algorithms in corporate practice. Furthermore, we provide an evaluative survey of recent scoring methods to aid future research.

## **Generalizations of Poisson Process in The Modeling of Random Processes Related to Road Accidents**

**Franciszek Grabski**

*Polish Naval Academy, Dept of Mathematics and Physics, Poland*

The stochastic processes theory provides concepts and theorems that allow to build probabilistic models concerning accidents. So called counting process can be applied for modelling the number of road, sea and railway accidents in the given time intervals. A crucial role in construction of the models plays a Poisson process and its generalizations. The nonhomogeneous Poisson process and corresponding nonhomogeneous compound Poisson process are applied for modeling the road accidents number and number of injured and killed people in Polish road. To estimate model parameters were used data coming from the annual reports of the Polish police.

**Keywords:** Nonhomogeneous Poisson process, nonhomogeneous compound Poisson process

## **Very Little or Nothing among Words Must Be Measured with Numbers**

**Larry D. Griffin**

*American University of Malta, Malta*

While accounting remains the oldest profession, words remain much older and have been lost among humankind's desire to bring humans to assume

responsibility for what some among them in the greed of their fellows only allows them to exert their own power for their own gain.

So those in power measure, and while measurements in numbers in quantity provide humans with one noetic of experience, nothing among their fine notations in numbers can ever alone replace the parallel noetic of the quality of words, the greater human experience, which remains practiced more often among all humans every day perhaps since the beginning of time. It takes quantity many millennia to even consider numbering time, and only after some back filling in the space of that time did the world produce anyone like an Einstein who could run the numbers and in words express quality in a way that anyone could even begin to understand the importance of the word or its relationship to time and space. While American education today under the duress of the power of the quantifiers who desire to manipulate numbers to meet political agendas of their own now has the power to use such numbers to undermine any quality of the word, those facilitators of the word among writers now refuse to cooperate.

### **Individual Control Charts for Ashes Content in Mineral Coal**

**Luís M. Grilo<sup>1,2,3,4</sup> and Teresa Oliveira<sup>5,6</sup>**

<sup>1</sup> *Deprt de Matemática, Instituto Politécnico de Tomar, Estrada da Serra - Quinta do Contador, 2300-313 Tomar, Portugal,* <sup>2</sup> *Centro de Matemática e Aplicações (CMA), FCT, UNL, Portugal,* <sup>3</sup> *Centro de Investigação em Cidades Inteligentes (Ci2), IPT, Portugal,* <sup>4</sup> *CIICESI / ESTG – P. Porto, Portugal,*

<sup>5</sup> *Deprt de Ciências e Tecnologia, Universidade Aberta, Rua da Escola Politécnica 141-147, 1269-001 Lisboa, Portugal,* <sup>6</sup> *Centro de Estatística e Aplicações, Universidade de Lisboa, Portugal.*

A variable measured in the labs of a mineral coal industry is the ashes content, which is an important variable to determine the calorific value of thermal coal. Based on the individual observations, available in this case study, Shewhart control charts are obtained computing the control limits using the average and also the median moving range (after analysing the assumptions of normality and the absence of autocorrelation of the observations). To detect small shifts in the mean, robust control limits are estimated using a non parametric method based on empirical quantiles with the bootstrap procedure, since the sample size is small. The Cumulative Sum (CUSUM) and Exponentially Weighted Moving Average (EWMA) control charts, considered sensitive to persistent assignable causes, are also obtained and compared, once the control limits should be made based on the required accuracy. Some statistics measures of process capability are calculated allowing to consider a capable process.

**Keywords:** Industrial process, robust control limits, Small sample, Statistical Process Control, Thermal coal.

## A Comparison of Covariate-based Prediction Methods for FIFA World Cups

Andreas Groll<sup>1</sup>, Gunther Schaubberger<sup>2</sup>

<sup>1</sup>Faculty of Statistics, Technical University Dortmund, Germany,

<sup>2</sup>Dept of Statistics, Ludwig-Maximilians-University Munich, Germany

Many approaches that analyze and predict the results of international matches in soccer are based on statistical models incorporating several potentially influential covariates with respect to a national team's success, such as the bookmakers' ratings or the FIFA ranking. Based on all matches from the four previous FIFA World Cups 2002-2014, we compare the most common *regression models* that are based on the teams' covariate information with regard to their predictive performances. Furthermore, an alternative modeling class is investigated, so-called *random forests* (Breimann, 2001).

Within the framework of Generalized Linear Models (GLMs), the most frequently used type of regression models in the literature is the *Poisson model*. It can easily be combined with different regularization methods such as penalization (see, e.g., Groll and Abedieh, 2013; Groll et al., 2015) or boosting (Groll et al., 2017). Moreover, we analyze different predictor structures, including team-specific ability parameters and extensions to smooth, non-linear effects for metric covariates, which also can be tackled by suitable boosting techniques (compare, e.g., Bühlmann and Hothorn, 2007).

*Random forests* can be seen as mixture between machine learning and statistical modeling and are known for their high predictive power. Here, we consider two different types of random forests depending on the choice of the response. One type of random forests tries to predict the precise numbers of goals while the other type considers the three match outcomes *win*, *draw* and *loss* using a special algorithm for ordinal response recently proposed by Hornung (2017).

For all these different modeling techniques the predictive performance with regard to several goodness-of-fit measures is compared. Based on the estimates of the best per-forming method all match outcomes of the FIFA World Cup 2018 in Russia are repeatedly simulated (1,000,000 times), resulting in winning probabilities for all participating national teams.

**Keywords:** Football, FIFA World Cups, Poisson regression, Random forests, Regularization.

### References

- [1].L. Breimann. Random Forests. *Machine Learning*, 45, 1, 5-32, 2001.

- [2].P. Bühlmann and T. Hothorn. Boosting Algorithms: Regularization, Prediction and Model Fitting (with Discussion). *Statistical Science*, 22, 4, 477-505, 2007.
- [3].A. Groll and J. Abedieh. Spain retains its title and sets a new record – generalized linear mixed models on European football championships. *Journal of Quantitative Analysis in Sports*, 9, 1, 51-66, 2013.
- [4].A. Groll, T. Kneib, A. Mayr, and G. Schaubberger. On the Dependency of Soccer Scores - A Sparse Bivariate Poisson Model for the UEFA European Football Championship 2016, In *Proceedings of MathSport International 2017 Conference*, 161-175, 2017.
- [5].A. Groll, G. Schaubberger, and G. Tutz. Prediction of major international soccer tournaments based on team-specific regularized Poisson regression: an application to the FIFA World Cup 2014. *Journal of Quantitative Analysis in Sports*, 11, 2, 97-115, 2015.
- [6].R. Hornung, Ordinal Forests. *Technical Report, Dept of Statistics, LMU*, 212, 2017.

## Robust Forecasting of Electricity Prices

**Luigi Grossi, Fany Nan**

*University of Verona, Dept of Economics, Italy*

In the present paper we suggest to use a version of threshold autoregressive models (SETARX) where parameters are estimated robustly to the presence of spikes. Differently from what has been done in the literature so far, we are not interested in modelling spikes, but we want to focus the attention on the influence that spikes can have on the estimated coefficients. If non robust estimators are applied, coefficient could be very badly biased and even non-spiky observations, which are the very large majority, could not be properly modeled and forecasted.

Moreover, we suggest a completely robust approach to modelling and forecasting electricity prices which embed robust estimation of a SETARX model, robust tests for unit root and nonlinear components and robust information criteria. Although we are aware of the limits of this class of models, threshold models represent a simple approach which takes into account the possible nonlinearity of electricity prices and allows the inclusion of external regressors to improve their forecasting performances. From the analysis of the existing literature, it is not clear the extent of the bias of robust estimators of the threshold with respect to LS estimator, how to choose the best weighting function and the forecasting performances of different weighting functions have never been compared. Moreover, robust estimators of regime switching processes are not implemented within the most popular software platforms among statisticians, such as Matlab and R. Grossi, L., & Nan, F. (2015) have started to address the above points through a Monte Carlo experiment which compares the performances of classical SETAR estimator and robust estimator using different weighting functions: the Tukey's function, the Huber's function and the polynomial function. The main result is that the bias in the threshold parameter estimator, which has been observed in previous works, decreases when Huber's and Polynomial weighting functions are applied, when the sample size increases and for complex

contamination patterns. However, when the features of the trajectories are more similar to what is observed on electricity markets, the polynomial function looks to be the best estimator. The simulation experiment has required the implementation of all the estimators (classical and robust) in R language resulting in a set of functions which hopefully will become a library soon.

The robust GM-estimator of SETAR processes based on the polynomial weights has been applied to forecast hourly day-ahead spot prices observed on the Italian market in the period 2013-2015. The long run trend has been estimated using a wavelet-based procedure and the stationarity of the detrended series has been verified through robust tests. The nonlinearity of the generating process has been robustly tested using a non-robust and robust tests. Finally the order of the SETAR model has been selected by a robust version of the Akaike Information Criteria.

Using prediction error statistics (MSE and MAE) and forecasting performance tests (Diebold and Mariano test and Model Confidence Set test), the nonlinear process SETAR(7,1) has revealed more effective than a linear AR(7) in predicting prices for year 2015, confirming the output of the robust test for nonlinearity. Besides the information set given by the past observations, several exogenous variables can be used to improve the forecasting performances of nonlinear models applied to electricity prices. Following recent literature, days-of-the-week dummy variables, predicted electricity demand and predicted wind power generation have been introduced as exogenous regressors in the SETAR(7,1) model on the Italian market.

The superiority of the forecasting performance of the robust on the LS estimator with exogenous regressor is overwhelming. The introduction of effective regressors, not only improve the forecasting power of the models, but the predictive ability of the robust estimator is significantly better than that of the LS estimator in more than 50% of the total cases.

## **Regional Impacts of Future Climate Change on Health and Labor in Brazil**

**Gilvan Guedes, Kenya Noronha, Aline Magalhaes, Edson Domingues, Kenia Souza, Flaviane Santiago, Debora Cardoso**  
*Cedeplar | UFMG, Brazil*

Global climate and environmental change have increased in the last decades. Increased health stress is one of the most alarming consequences of these changes. Although many studies have tried to estimate the direct and indirect consequences of a warmer and drier environment for the economy, both at a global and local scale, a smaller number of studies have addressed the mid and long term health



implications of these changes at a regional level. Building on their previous work, this study takes a multi-stage approach to estimate the climate-related consequences on cardiovascular/respiratory and infectious/vector-borne diseases, morbi/mortality, and labor supply in Brazil. Combining Spatial Bayes Smoothing, Spatial Econometrics, Global Burden of Disease data, and a Regional Computable General Equilibrium model, this study estimates the future development of climate-sensitive health disorders, their implications for morbi-mortality, and the consequences for labor supply and productivity for the Brazilian states and regions from 2010 to 2040. Our results suggest that partial effects of climate change on health and labor supply is higher than the total impact (from general equilibrium estimates). Increased morbi-mortality and labor loss would be higher for vector-borne and infectious than for non-communicable diseases, and mostly concentrated in less developed regions of the country.

**Keywords:** Climate Change, Health, Labor Supply, Brazil, Computable General Equilibrium Model, Regional Impacts

## Long Memory Process under Strong Mixing Noises

**Soraya Haddad, Karima Belaïde**

*Dept of Mathematics, Applied Mathematics Laboratory,  
University of Bejaia, Algeria*

This work is devoted to the study of the behavior of long memory process, we are going to study the behavior of the probabilistic properties of fractional autoregressive long memory model in the case of strong mixing noises. Firstly, we are going to give some results about the autoregressive function and long memory process in the case of mixing noises. The underlying goal is to provide that the mixing noises influence the behavior of the process.

**Keywords:** Long memory, strong mixing, independence, fractional autoregressive process

## Exploring the Dynamics of Business Survey Data with Markov Models

**W. Hölzl<sup>1</sup>, S. Kaniovski<sup>1</sup>, Y.M. Kaniovski<sup>2</sup>**

*<sup>1</sup>Austrian Institute of Economic Research (WIFO), Austria, <sup>2</sup>Faculty of Economics and Management, Free University of Bozen-Bolzano, Italy*

Business tendency surveys are widely used for monitoring economic activity. They provide a timely feedback on the current business conditions

and outlook. We analyze the distribution of the three-way responses regarding the current business climate and production and identify the unobserved macroeconomic factors behind them. The quarterly counts characterize the Austrian economy in the period of time between 1991 and 2017.

Aggregate and micro-founded models are considered. The former endogenously identify two macroeconomic regimes: upturn and downturn. Their dynamics is modeled with a regime-switching matrix. A micro-founded model generates responses dependent across the firms: unobserved common macroeconomic factors, favorable or adverse, imply more frequent optimistic or, respectively, pessimistic opinions. The corresponding conditional transition probabilities are defined according to a coupling scheme. Extensions addressing a sector dimension and a dynamics of the common tendencies are considered. This macroeconomic dynamics is modeled with a hidden Markov chain.

**Keywords:** business tendency surveys, business cycle, coupled Markov chain, multinomial distribution

## **A Latent Class Analysis of the Profile and Satisfaction of Users of the Social Care Units in Greece**

**Konstantina Kakava, Dimitrios Kalamaras, Aglaia Kalamatianou**

*Dept of Sociology, Panteion University of Social and Political Sciences, Greece*

Latent class analysis can be considered as a special case of model-based clustering of multivariate discrete data. The idea is that each observation comes from one of a number of classes, groups or subpopulations, with its own probability distribution. The overall population thus follows a finite mixture model. In this study, we used the clustering approach based on the model for grouping and detecting heterogeneities regarding the satisfaction of users of the elderly social care units in Greece from the services provided taking into account the socio-demographic profile of these users. The study data comes from a targeted survey and findings from the latent class analysis are also compared with those provided using other methods for analysis of categorical data. The data analyzed using R software. Overall the results of the study could be considered as a good indicator for assessing the effectiveness of elderly social care units.

**Keywords:** Latent class analysis, satisfaction with services provided from elderly social care units

## On Generalized Linear Mixed Modeling of Epidemic Prone Syndromes in Biosurveillance

Emmanouil-Nektarios Kalligeris<sup>1</sup>, Alex Karagrigoriou<sup>1</sup>,  
Christina Parpoula<sup>1</sup>, Angeliki Lambrou<sup>2</sup>

<sup>1</sup>*Lab of Statistics and Data Analysis, Dept of Mathematics, University of the Aegean, Greece,* <sup>2</sup>*Hellenic Center for Disease Control and Prevention, Greece*

A rich class of traditional statistical models and methods is currently available for the early detection of epidemic activity in biosurveillance systems. Real time surveillance though is often difficult to be fully achieved because of the seasonality involved in the series. Indeed, whenever the correlation structure of a series depends on the season, the time series involved fails to reach stationarity with all the associated modeling consequences. In such situations, a useful class of models is that of periodic autoregressive moving average (PARMA) models (see e.g. Anderson et. al, 2013; Jones and Brelsford (1967); Pagano (1978); Tesfaye et. al. (2006)), which are extensions of commonly used ARMA models to allow parameters that depend on season. In this work, general as well as special PARMA models are considered, and via likelihood and model selection identification techniques, the optimal model for a weekly influenza-like syndrome data set is selected. Climatological and weather covariates associated with influenza-like syndrome are also incorporated into the analysis. The resulting model succeeds in identifying the epidemic wave, and is found to outperform typical forecasting models.

**Keywords:** data analysis, periodic autoregressive moving average models, influenza, biosurveillance, model selection.

### References

- [1]. P. L. Anderson, M. M. Meerschaert, and K. Zhang. Forecasting with prediction intervals for periodic autoregressive moving average models, *Journal of Time Series Analysis*, 34(2):187–193, 2013.
- [2]. R. H. Jones and W. M. Brelsford. Time series with periodic structure. *Biometrika*, 54:403–408, 1967.
- [3]. Y. Tesfaye, M. M. Meerschaert, and P. L. Anderson. Identification of periodic autoregressive moving average models and their application to the modeling of river flows, *Water Resources Research*, 42, W01419, 2006.
- [4]. M. Pagano. On periodic and multiple autoregressions, *Annals of Statistics*, 6(6):1310–1317, 1978.

## Distributionally Robust Optimization via Optimal Transport and its Applications in Machine Learning

**Yang Kang**

*Columbia University, USA*

Optimal mass transportation is a powerful tool in the arsenal of many quantitative disciplines, with well-documented applications spanning a wide range of areas, including, operations research, economics and image analysis. In this talk, we focus on data-driven distributionally robust optimization, that is, a class of perfect-information games in which an optimizer selects an action and adversary chooses a model within a region around a baseline distribution, which we often take to be an empirical measure. We show how many machine learning algorithms can be retrieved as special cases of this type of formulation. We establish connections to regularized portfolio optimization strategies that are common in practice. These connections provide a rich intuition which allows interpreting various regularization parameters which are typically chosen in practice via cross-validation, but owing to this intuition, we are able to define a reasonable optimization criterion for choosing regularization parameters via pivotal statistics, thereby avoiding time-consuming cross-validation.

*(This talk is based on joint work with Jose Blanchet, Karthyek Murthy, and Fan Zhang).*

**Keywords:** Distributionally Robust Optimization, Optimal Transport

## Change Point Detection in Activity Monitoring using Multivariate Statistical techniques

**Naveed Khan, Sally McClean, Shuai Zhang, Chris Nugent**

*School of Computing, Ulster University, UK*

The monitoring of human activities using simple body worn sensors is an important and emerging area of research in machine learning. These sensors can capture a large amount of data in a short period of time in a relatively unobtrusive manner. The sensor data might have different transitions to be used for analysing diverse user activities. Therefore, change point detection can be used to classify the transition from one underlying distribution to another. Automatic and accurate change point detection is not only used for detecting different events, however, but can also be used for generating real world datasets and also responding to changes in patient vital signs in critical situations.

In this paper, we use multivariate exponentially weighted moving averages (MEWMA) for online change point detection. Additionally, genetic

algorithm (GA) and particle swarm optimization (PSO) is used to automatically identify an optimal parameter set by maximizing the F-measure. Furthermore, we evaluate our approach against multivariate cumulative sum (MCUSUM) from the state-of-the-art in terms of different metric measures such as accuracy, precision, sensitivity, G-means and F-measure. Results have been evaluated using real data set of 10 users performing 9 distinct activities for a total period of 35 minutes for accurate change point detection. Moreover, the results show that the proposed scheme outperforms significantly compared with the benchmark scheme by achieving high accuracy, precision, sensitivity, G-means and F-measure while utilizing minimum computational cost for accurate change point detection in multivariate data analysis.

**Keywords:** Activity Monitoring, Particle Swarm Optimization, Genetic Algorithm, Multivariate Exponentially Weighted Moving Average, Multivariate Cumulative SUM.

## Ruin Probability in Multivariate Collective Risk Model with Dependent and Heavy-Tailed Components

**Yury Khokhlov**

*Dept of Mathematical Statistics in Lomonosov Moscow State University, Russia*

We propose the new variant of the collective risk model which is the development of the model from Ivanova and Khokhlov (2007). We consider the problem of ruin probability estimation in such model for the case of large initial capital. To solve this problem we use the methods from Konstantinides and Li (2017) but consider some different model. Also we consider two different approaches for the definition of the notion of ruin probability in multivariate case.

**Keywords:** multivariate collective risk model, heavy-tailed distributions, ruin probability

**Acknowledgements:** *This research is supported by RSCF, project 18-11-00155.*

### References

- [1].N.L. Ivanova and Yu.S. Khokhlov, Multidimensional collective risk model. Journal of Mathematical Sciences, Volume 146, Number 4, October 2007, p. 6000-6007.
- [2].D.J. Konstantinides, J.Z. Li. Ruin Probabilities for a Multidimensional Renewal Risk Model with Multivariate Regularly Varying Claims. – Analytical and Computational Methods in Probability Theory and its Applications. – Proceedings of the International Conference. – 23-27 October 2017, Moscow, Russia, p.p. 446-452.

## **A Nurse Dyad Model for Collaborative Action Planning and Goal Attainment to Support Patient Self-Management of Stroke Risk Factors**

**Barbara Kimmel MS., MSc, MPH <sup>2,3</sup>; Jane Anderson PhD, RN, FNP-BC, FAANP <sup>1-2</sup>, Paul Swank, PhDc and Ross Shegog, PhD<sup>3</sup>**

*<sup>1</sup>Houston VA HSR&D Center for Innovations in Quality, Effectiveness and Safety, Houston, TX, <sup>2</sup>Department of Neurology, Baylor College of Medicine, Houston, TX*

*<sup>3</sup>University of Texas, School of Public Health, Department of Health Prevention and Behavioral Science, Houston, TX*

**Background and Purpose:** Patient self-management (SM) of stroke risk factors is critical to prevent second stroke. Development of an action plan to reach desired health behavior goals is well established for chronic disease SM. In collaborative planning, patients identify a goal and then a healthcare provider helps develop action plan to attain the goal and reduce stroke risk factors. We pilot-tested a model where nurses in hospital stroke ward and in primary care worked as dyads to support stroke patients in setting goals. We conducted a pilot of a nurse dyad model to determine feasibility of coaching patients in goal-setting and the Goal Attainment Measure for Stroke (GAM-S).

**Innovation:** Nurses have been identified as front line healthcare providers for patient education and self-management support. The nurse dyad model for collaborative action planning may help stroke survivors achieve and sustain goals for optimal management of stroke risk factors.

**Methods:** A pre-experimental, pre/post design was applied. Ten registered nurses dyads and 44 stroke patients were recruited from neurology at the Michael E. DeBakey Veterans Administration Medical Center in Houston, Texas. Nurses from the hospital unit delivered one-on-one educational sessions to patients admitted with stroke and set goals and action plans in patients' medical records. Nurses in primary care clinic called patients 2-weeks after the hospital discharge to evaluate goal attainment, which was documented and scored using GAM-S. The GAM-S components are scored as, 2 = fully attained, 1 = partially attained, and 0 = not attained.

**Results:** 44 patients were enrolled in the study. They were 42 males (95%) and 2 female (5%). Mean patients' age was 67 years old with 23 White (52%), 10 African American (23.5%), 10 Hispanics (23.5%), 1 Other (1%). GAM-S completion rate: out of 44 patients enrolled, 42 completed GAM-S (95% completion). Patients set up action plans to address modifiable stroke risk factors control such as diet, exercise, reduction in cigarettes smoking and in alcohol consumption. In addition, some patients have decided to improve their blood pressure and diabetes medication compliance. Specifically, 12 patients (29%) decided to increase their physical activity, 17 patients (40 %) wanted to make healthy improvements in their eating habits. Eight patients (19%) were willing to

reduce smoking and 3 patients (7%) decided to reduce alcohol intake. Finally, two patients (5%) made improvements in their medication compliance. Overall,

52% of patients achieved maximum total score equal to 8 indicating that these patients fully achieved each of their specific section of the Action Plan set up in an individual domain. The remaining score values ranged from 0 to 7 with the total of 5 patients (11%) not achieving any of their previously set up goals. The mean score for goal attainment was 5.9.

Conclusions: Implementation of a nurse dyad model for collaborative action planning is feasible and shows promise for engaging patients in self-management of stroke risk factors control. Findings from this study will help design a self-management support intervention for stroke survivors using a nurse dyad delivery model.

**Keywords:** self-management, stroke, stroke risk factors, GAM-S complexion rate

### References

- [1]. Anderson, Jane A, Kyler Goldwin, Jason Saleem, Scott Russell, Jashua Robinson, Adam Netto and Barbara Kimmel: Accessibility, Usability and Feasibility of a Web-based Clinical Decision Support Tool to Enhance Provider-Patient Communication around Self-management TO Prevent (STOP) Stroke. *Health Informatics Journal*, 2012.
- [2]. Anderson, J.A., Godwin, K.M., Petersen, N.J., Willson, P.C. (2013). A pilot test of videoconferencing to improve access to a stroke risk-reduction programme for Veterans. *Journal of Telemedicine and Telecare*; 19(3), 135 – 141.
- [3]. Damush, T.M., Myers, L., Anderson, J.A., Yu, Z., Ofner, S, Nicholas, G., Kimmel, B., Schmid, A., Kent, T. & Williams, L.S The Effect of a Locally Adapted, Secondary Stroke Risk Factor Self-Management Program on Medication Adherence among Veterans with Stroke/TIA. (2015).. *Translational Behavioral Medicine*. Published October 26, doi: 10.1007/s13142-015-0348-6
- [4]. Jones, F. (2006). Strategies to enhance chronic disease self-management: how can we apply this to stroke?. *Disability and rehabilitation*, 28(13-14), 841-847.
- [5]. Latham, G. P., & Locke, E. A. (1991). Self-regulation through goal setting. *Organizational behavior and human decision processes*, 50(2), 212-247.
- [6]. Lorig K, Holman H. (2003) Self-management education: History, definition, outcomes, and mechanisms. *Ann Behav Med* 26: 1–7. PMID: 12867348
- [7]. Shegog, R., Begley, C. E., Harding, A., Dubinsky, S., Goldsmith, C., Hope, O., & Newmark, M. (2013). Description and feasibility of MINDSET: a clinic decision aid for epilepsy self-management. *Epilepsy & Behavior*, 29(3), 527-536.
- [8]. Shegog, R., Begley, C., Iyagba, B., Harding, A., Dubinsky, S., & Newmark, M. (2011). MINDSET: the development and testing of a clinic-based self-management decision support system. In *American Epilepsy Society Annual Meeting, Special Interest Group Session of PEC*.

## Testing Covariance Structures: Non-Normal Approach

Tõnu Kollo, Marju Valge

*Institute of Mathematics and Statistics, University of Tartu, Estonia*

Test-statistics for sphericity and uncorrelatedness hypothesis on covariance structure are examined for non-normal populations. We use

test-statistics based on trace functions. Existence of the fourth order moments is assumed. In a special case the results are simplified when all the fourth order moments are equal. Taylor expansions of the test statistics have been derived, asymptotic normality and chi-square distributions have been established and their behaviour examined in the situation when both, sample size  $n$  and number of variables  $p$  tend to infinity,  $p/n < 1$ . Parallel to theoretical study a simulation experiment is carried out to investigate empirically speed of convergence to asymptotic distributions depending on sample size, number of variables and parameters of the population distribution.

**Keywords:** asymptotic normality, chi-square distribution, sphericity test, uncorrelated-ness test

## Ruin Probabilities for a Double Renewal Risk Model with Frequent Premium Arrivals

Dimitrios G. Konstantinides

*Dept of Mathematics, University of the Aegean, Greece*

In this paper a double renewal risk model is studied. The claims represent i.i.d. sequence of random variables and the premiums represent another sequence of random variables with lower extended negative dependence. The corresponding two arrival processes have different intensities, which correspond to consideration of frequent arrivals of premiums. The ruin probability over a finite horizon is asymptotically estimated when the initial capital tends to infinity.

**Keywords:** asymptotics, premium process, claim process, multivariate regular variation.

### References

- [1]. Bingham, N.H., Goldie, C.M., Teugels, J.L. (1987) Regular Variation. Cambridge University Press, New York.
- [2]. Chen, Y., Chen, A., Ng, K.W. (2010) The strong law of large numbers for extended negatively dependent random variables. J. Appl. Probab., 46, 908-922.
- [3]. Gao, Q., Wang, Y. (2010) Randomly weighted sums with dominated varying-tailed increments and application to risk theory. J. Korean Stat. Society 39, 305-314.
- [4]. Nelsen, R.B. (2006) An Introduction to Copulas. Springer, Berlin.
- [5]. Robert, C. Y., Segers, J. H. (2008) Tails of random sums of a heavy-tailed number of light-tailed terms. Insurance: Math. and Econom., 43, 85-92.
- [6]. Ross, S. M. (1983) Stochastic Processes. Wiley, New York.
- [7]. Schmidli, H. (1999) Compound sums and subexponentiality. Bernoulli, 5(6) (1999), 999-1012.
- [8]. Tang Q. H., Tsitsiashvili G. (2003) Precise estimates for the ruin probability in finite horizon in a discrete-time model with heavy-tailed insurance and financial risks. Stochastic Process. Appl., 108, 299-325.
- [9]. Tang Q. H., Tsitsiashvili G. (2003) Randomly weighted sums of subexponential random variables with application to ruin theory. Extremes, 6, 171-188.



## The Impact of the Current Economic Crisis on Mortality in Greece

**Anastasia Kostaki<sup>1</sup>, Konstantinos Zafeiris<sup>2</sup>, Byron Kotzamanis<sup>3</sup>**

<sup>1</sup>*Dept of Statistics, Athens University of Economics and Business, Athens, Greece*

<sup>2</sup>*Laboratory of P. Anthropology, Dept of History and Ethnology, Democritus University of Thrace, Greece,*

<sup>3</sup>*Dept of Urban Planning and Regional Development, University of Thessaly, Volos, Greece*

The relationship between economic conditions and mortality of a population, unlike other demographic phenomena as fertility and migration, is not as clear as expected. A reasonable assumption is that in conditions of economic recession, which equates to increased unemployment, falling disposable income and reduced access to health care systems, the mortality of a population tends to increase, while in conditions of economic boom, it tends to decrease. However, this expected negative relationship between economic conditions and mortality, is not always verified by empirical research but in many cases these studies show contradictory and conflicting results, which causes intensive debates and controversy in the scientific community. The question that causes strong controversy in the literature, are summarized as follows: What is the impact of the circularity of the financial system on mortality and health? Based on data from Greece mortality trends and the impact of economic crisis on them will be analyzed in this paper.

**Keywords:** Greece, economic crisis, mortality

## Orlicz Duality in Stochastic Processes and Risk Measures' Applications

**Christos E. Kountzakis, Dimitrios G. Konstantinides**

*University of the Aegean, Greece*

We transfer some extensions of  $L_p$ -duality concerning the associated monetary risk measures' representation, to a more general theory of Orlicz spaces' duality of stochastic processes. The main motivation for the present paper is to find classes of Young functions for which the risk measure -as an amount of capital- is finite for a given stochastic process, as a vector of an Orlicz space.

## A New Bivariate Distribution with Gamma and Beta Marginals

**Markos V. Koutras, Theodora K. Kokkinioti**

*Dept of Statistics and Insurance Science, School of Finance and Statistics,  
University of Piraeus, Greece*

In multivariate statistical analysis, special research interest has been drawn in the development of techniques for the construction of new bivariate distributions with given marginals. One of the most popular methods to achieve this is the trivariate reduction method in which two appropriate transformations of a set of independent variables is considered so that the two new variables inherit some type of dependence. In the majority of publications that exploit this method, the two marginals are of the same type, for example several bivariate gamma distributions have been introduced with their marginals being both gamma.

In this work, we made use of the trivariate reduction method in order to introduce a new bivariate distribution, whose marginals are of different type, specifically one of them is a Gamma distribution and the other is a Beta distribution. Such a distribution may be proved quite useful in many applied areas, when one of the variables of interest is bounded (e.g. a proportion) and the other one is unbounded (e.g. lifetime).

**Keywords:** bivariate distribution, trivariate reduction method, stochastic dependence

## Compound Sooner Waiting Time Problems in Multistate Trials

**Markos V. Koutras, Ioannis S. Triantafyllou**

*University of Piraeus, Greece, University of Thessaly, Greece*

In many applications, one may face the problem of analyzing the functionality of units or systems that are subject to two different types of failures. These cases can be modeled by considering sequences of trials with three possible outcomes in each one (say F1: failure of type I, F2: failure of type II, S: success) and looking at the waiting time  $T$  until the first occurrence of an F1-run of length  $k$  or an F2-run of length  $r$  ( $k \leq r$ ). If  $Y_t$  denotes the interarrival time between the  $(t-1)$ th and  $t$ -th trial, then the compound random variable  $S$  is potentially useful for modeling the lifetime of a system under the above structure. In the present article, we develop some results for the distribution of the random variable  $S$  in the case where  $Y_t$ 's are discrete random variables. More precisely, the probability

generating function and the moments of  $S$  are studied, while the problem of evaluating its probability mass function is also addressed.

## Estimating Probabilities of Default for Portfolios with Low or Zero Defaults

**Vasileios M. Koutras**

*Dept of Statistics and Insurance Science, School of Finance and Statistics,  
University of Piraeus, Greece*

In finance, it is widely recognized that the efficient estimation of probabilities of default (PD) per borrower is of utmost importance for it provides the main ingredient of credit risk modelling and managing techniques. The accuracy of the PD estimations determines the quality of the results of credit risk models.

Among the most common obstacles that make the PD estimation procedure intractable or difficult to apply can be the low number of defaults, especially in the more creditworthy rating grades, which may enjoy many years without any defaults. Even if some defaults occur in a given year, the observed default rates might exhibit a high degree of volatility due to the relatively low number of cases (i.e. loans) in that grade. Because of the significant practical importance of the aforementioned problem, it has attracted considerable recent research attention. Several techniques have been suggested for tackling the problem of the lack of sufficient data to carry out the typical statistical analysis. For example, the use of upper confidence bounds with a prespecified confidence level and the engagement of Bayesian analysis, provide the required means for resolving the problem of low or zero defaults in the portfolio under study. In the present work we present several techniques that can be used in the estimation of low default probabilities and carry out a detailed comparison of these techniques through a simulation study. The empirical findings may be of major importance to the various parties involved in the financial system (e.g. Fed, ECB, Monetary Financial Institutions). More specifically, the aforementioned findings could be applied in the assessment of the creditworthiness of various types of loan portfolios (e.g. Project Finance, Real Estate etc.) and/or the surveillance of the credit quality of these portfolios by the corresponding supervisory authorities (e.g. FDIC, EBA etc.).

**Keywords:** Financial risk management; risk modeling; low default probabilities; upper confidence bounds; Bayesian analysis.

**Acknowledgements:** *The work presented in this paper has been partially funded by National Matching Funds 2016-2017 of the Greek Government, and more specifically by the General Secretariat for Research and Technology (GSRT), related to EU project "ISMPH: Inference for a Semi-Markov Process using Hazards Specification" (GA No 329128).*

## Optimal Maintenance Policies of a Two Unit Multi-State Deteriorating System with Imperfect Switch

Vasilis P. Koutras<sup>1</sup>, Sonia Malefaki<sup>2</sup>, Agapios N. Platis<sup>1</sup>

<sup>1</sup>Dept of Financial and Management Engineering, School of Engineering, University of the Aegean, Greece, <sup>2</sup>Dept of Mechanical Engineering & Aeronautics, University of Patras, Patras, Greece

In this work, a two-unit multi-state deteriorating system under preventive maintenance and imperfect switch among units is considered. The system consists of one operating and one standby unit, in cold standby mode. System control is switched to the standby unit when the operational unit experiences a failure. The switching automated mechanism though, can experience failures either due to often use that incur aging and degradation effects, or even due to extended periods of being idle resulting in failing to operate when it is needed. In order to prevent or avoid operational unit failures due to deterioration, maintenance actions (either minimal or major) are proposed to be implemented to the operational unit. However, operational unit maintenance can be also imperfect, restoring the unit to a worse degraded state, or even to a total failure state, mainly due to external factors. The operational unit is periodically inspected in order to distinguish if any maintenance action needs to be triggered. The main aim of this work consists in modeling the behavior of the aforementioned two-unit system under a Markov framework and in examining how unit inspection intervals, as well as switching mechanism success probability, affect the entire system dependability and performance. Towards this direction, system asymptotic availability as well as total expected operational cost are derived and evaluated through experimental numerical examples. The endmost purpose of our work is to determine an optimal policy for unit inspection that improves system dependability and performance measures.

## Modeling and Solving A Dynamic Winner Determination Problem Under Three Reals Objectives

Larbi Asli, Méziane Aider

Faculté des Sciences Exactes Targa Ouzamour, Algeria

In this work, we propose a mathematical formulation model for the dynamic three objectives winner determination problem (DWDP), in combinatorial auctions. This model is developed for the purpose of implementing the web, the model reflects the English auction mechanism.

We have generalized the algorithm of resolution that we have previously designed for the bi-objective WDP. It is based on the temporal analysis of the bid status in order to avoid to the maximum to the combinatorial resolution. We use the evolutionary algorithm if need be, according to the state of the bid. The algorithm gives at every time of the auction period, a temporary list of winners, articles sold and the temporary gain, and even the delivery time. We use simulated data to test the adequacy and the effectiveness of the mechanism.

**Keywords:** English combinatorial auction, three objective, WDP, bi-objective WDP, evolutionary algorithm, temporary winners list, final winner, winner determination problem

## Quality of Statistics on Cause of Death

**Tae Rim Lee**

Dept. of Information & Statistics, Korea National Open University, Rep. of Korea

In order to improve the quality of statistics in Korean health sector, the basic goal is to evaluate the quality of statistical data on 23 of the 50 health statistics approved by the National Statistical Office. However, rather than diagnosing the 23 species of health statistics individually, this study takes the overall position of these statistics in the context of the current health environment in Korea, and achieves the health policy goal of improving health and improving quality of life. The main content of the study is how useful cause of death information is provided and what efforts are needed to provide better information.

Statistics on causes of death are statistics based on death reports and death certificates obtained through population trend surveys. They are the basis for establishing and evaluating the health and welfare policies of the citizens by analyzing the characteristics of the deaths.

The purpose of this study is to collect and reflect the user requirements through user group meetings and to collect data on the cause of death statistics and the improvement measures were reconciled with the opinion of the field which calculated actual cause of death, and major improvements were derived.

The quality report is based on six dimensional diagnostic procedures [relevance (5.0), accuracy (4.6), timeliness / timeliness (5.0), comparability (5.0), consistency (4.0), accessibility / clarity (3.4). The 10 steps process was screened ① Summary (5.0), ② Statistics use (5.0), ③ Survey design (4.7), ④ Data collection (4.9), ⑤ administrative data (not applicable), ⑥ data processing (4.7), ⑦ statistical estimation (5.0), ⑧ statistical publication (4.2) ⑨ statistics fundamental(3.5) ⑩ **References:** (consisting of additional scores only)] The statistical diagnosis was

performed according to the diagnosis check list. The overall evaluation score for the quality was 96.5, which was very good.

## Single-year Estimates in Developing Countries Mortality Database (DCMD)

Nan Li<sup>1</sup>, Hong Mi<sup>2</sup>

<sup>1</sup>Population Division, Dept of Economic and Social Affairs, United Nations, USA,

<sup>2</sup>School of Public Affairs, Zhejiang University, P. R. China

In the earlier phases of Developing Countries Mortality Database (DCMD), the estimates of old-age mortality (45Q60 or Qo) and life tables used only census population at old ages. These estimates referred to the periods between successive censuses. This paper describes the method of converting period estimates into single-year estimates, using the relationship between period and single-year estimates and, when available, utilizing also the data on old age deaths in census years collected from death registration or census.

**Note:** *The views expressed in this paper are those of the author and do not necessarily reflect those of the United Nations.*

**Acknowledgement:** *The work on this paper was supported by the Nature Science Foundation of China (NSFC) project (NO.71490732), NSFC project (NO.71490733), Zhejiang Social Science Planning Project Key Program (NO.17NDJC029Z), and Nature Science Foundation of Zhejiang project (NO. LZ13G030001)*

## CLUSTATIS: A Cluster Analysis of Multiblock Datasets

Fabien Llobell<sup>1,2</sup>, Véronique Cariou<sup>1</sup>, Evelyne Vigneau<sup>1</sup>,  
Amaury Labenne<sup>2</sup>, El Mostafa Qannari<sup>1</sup>

<sup>1</sup>StatSC, ONIRIS, INRA. Nantes, France, <sup>2</sup>Addinsoft, XLSTAT, Paris, France

CLUSTATIS is a general procedure of cluster analysis of a collection of datasets. It is based on the optimization of a criterion and consists of a hierarchical cluster analysis and an iterative algorithm akin to K-means. Its interest is discussed and illustrated in sensory and preference studies.

**Keywords:** Cluster analysis, STATIS, Multiblock datasets, Sensory analysis.

## Some Results for the Moments of the Overshoot in a Random Walk with Positive Drift

**Losidis Sotirios, Konstadinos Politis**

*Dept of Statistics and Insurance Science, Greece*

Hereby we study the moments of the overshoot in a random walk with positive drift. We give a new formula for the right tail of the overshoot and discuss the monotonicity of its variance under the assumption that the distribution of the ladder height is IMRL. We finally study the asymptotic covariance between the overshoot and the first passage time.

**Keywords:** renewal function; renewal density; forward recurrence time; IMRL; HNBUE; HNWUE.

## Analysis of Sex-Age-Cause Specific Mortality in European Countries

**Aleša Lotrič Dolinar<sup>1</sup>, Simona Korenjak-Černe<sup>1</sup>, Jože Sambt<sup>1</sup>,  
Filipe Afonso<sup>2</sup>, Edwin Diday<sup>3</sup>**

*<sup>1</sup>Academic Unit for Mathematics, Statistics and Operational Research, Faculty of Economics, Kardeljeva ploščad 17, University of Ljubljana, Ljubljana, Slovenia,*

*<sup>2</sup>Symbad – Le Symbolic Data Lab, France, <sup>3</sup>Université Paris-Dauphine, France*

The mortality level and the structure of deaths by cause of death reflect health conditions that people face in different countries. Positioning of a country by those two dimensions is important information for adequate health policy. Deaths from some causes of death are easier to prevent than from others, which should be considered when allocating resources across prevention (and/or cure) programmes.

We analyse Eurostat data on sex-age-cause specific mortality for 32 European countries in 2014. We use classical analysis, but to preserve distribution over death causes in each sex-age combination we also use symbolic data analysis that enables richer data descriptions and therefore retains more information. Clustering results show that groups of countries are related to the geographical position of countries. We study the importance of both, mortality level and death cause structure, in country grouping. We also identify those sex-age-cause combinations that discriminate the most between individual countries as well as between resulting clusters. Another useful insight into countries' health situations is obtained through correlations between sex-age-cause specific mortalities, and between sex-age-cause-specific mortalities and mortality levels.

**Keywords:** mortality, causes of death, clustering, symbolic data analysis.

## Censored Kalman Smoother

**Kostas Loumponias, George Tsaklidis**

*Dept of Mathematics, Statistics and Operations Research Section, Aristotle University of Thessaloniki, 54124 Thessaloniki, Greece*

In this paper we establish the censored Kalman smoother, i.e. the Kalman smoother when the observations of the process are censored (partially known). The standard Kalman smoother provides optimal estimation of an unknown state vector, at time  $t$ , when all observations of a single time series are available. One of the basic assumptions of the standard Kalman smoother is that the observations are normally distributed; therefore, this filter cannot handle the censored observations optimally. Thus, in order to deal with this burden, we inject the “censored” normal distribution in the Kalman smoother process.

## Critical Points of the Vandermonde Determinant on Spheres Defined by Various Norms

**Karl Lundengård<sup>1</sup>, Asaph Keikara Muhumuza<sup>1,2</sup>, Jonas Österberg<sup>1</sup>, Sergei Silvestrov<sup>1</sup>, John Mango<sup>3</sup>, Godwin Kakuba<sup>3</sup>**

*<sup>1</sup>Division of Applied Mathematics, The School of Education, Culture and Communication, Mälardalen University, Sweden, <sup>2</sup>Dept of Mathematics, Busitema University, Uganda, <sup>3</sup>Dept of Mathematics, College of Natural Sciences, Makerere University, Uganda*

Finding the critical points of the Vandermonde determinant of a surface have applications for optimal conjugation of points for polynomial interpolation, curve-fitting and optimal experiment design. Here we examine the critical points on spheres defined by various norms and see how their coordinates can be expressed as roots of polynomials. The properties of these polynomials will be examined with the goal to and efficient ways of solving them and visualizing the results.

**Keywords:** Vandermonde determinant, orthogonal polynomials



## **Application of a Power-Exponential Function Based Model to Mortality Rates Forecasting**

**Karl Lundengård, Belinda Strass, Andromachi Boulougari,  
Milica Rancic, Sergei Silvestrov**

*Division of Applied Mathematics, The School of Education, Culture and  
Communication, Mälardalen University, Sweden*

Mortality rates of living organisms or equipment are modelled in different ways. Variation of mortality over a life span has different characteristics that put constraints and requirements on a model developed to represent it. A well-know problem that complicates modelling of human mortality rates is the "accident hump" occurring in early adulthood. The mortality rate model based on power-exponential functions, previously proposed by the authors, behaves as expected in that life period. Here, it will be compared to other models usually applied in practice and to empirical data. Models will be fitted to known data of measured death rates from many different countries using numerical techniques for curve-fitting with the non-linear least squares method. The properties of the model with respect to quality of fit and usefulness in applications such as insurance pricing or forecasting will be discussed.

**Keywords:** Mortality rates modelling, power-exponential function, non-linear curve fitting.

## **Change Point Analysis on Seismic Sequences of Gulf of Corinth**

**Lykou Rodi, George Tsaklidis, Elephtheria Papadimitriou**

*Aristotle University of Thessaloniki, Greece*

The Gulf of Corinth is of special seismological interest because of its intense seismic activity. Seismicity is strongly clustered and therefore change point analysis on such data seems suitable. Thus, multiple change point analysis is applied to recent seismic sequences. This procedure employs both online and offline methods providing the study with different perspectives. In parallel, results from univariate approaches are compared with those of multivariate ones for further inference on the spatiotemporal properties of the seismicity in the Gulf.

**Keywords:** Applied statistics, change point analysis, earthquake sequences, Gulf of Corinth

## Changes in Mortality in the Context of Economic and Social Development – Example of Selected European Countries

**Justyna Majewska, Grażyna Trzpiot**

*Dept of Demography and Economic Statistics, University of Economics in Katowice, Poland*

Population ageing is a natural and inevitable process in each country with a highly developed economy. Changes in mortality and disease profiles in conjunction with progress in health and socio-economic development allow for analysis of an epidemiological transition, and in broader notion – health transition. The aim of this article is to analyze the health status and mortality causes among people aged 65 years and over in selected European countries that are in different stages of health transition. Selected health measures and method of decomposing differences in expected life expectancy are used. A subjective selection of variables important from the point of view of improving the health of elderly people was made.

**Keywords:** health transition, ageing, life expectancy, DALYs

## Advanced Monte Carlo Pricing of European Options in a Market Model with Two Stochastic Volatilities

**Anatoliy Malyarenko<sup>1</sup>, Ying Ni<sup>1</sup>, Betuel Canhanga<sup>2</sup>,  
Sergei Silvestrov<sup>1</sup>**

*<sup>1</sup>Division of Applied Mathematics, School of Education, Culture and Communication, Mälardalen University, Sweden, <sup>2</sup>Faculty of Sciences, Dept of Mathematics and Computer Sciences, Eduardo Mondlane University, Mozambique*

We consider a market model with four correlated factors and two stochastic volatilities, one of which is rapid-changing, while another one is slow-changing in time. Advanced Monte Carlo methods based on the theory of cubature in Wiener space, are used to find the no-arbitrage price of the European call option in the above model.

**Keywords:** Stochastic volatility, market model, Monte Carlo method.

## **Optimal Scaling and Contingency Tables Reveal the Mismatch between Patients' Attitude and Perception Towards their Asthma Medications and Complaints during the I-MUR Service Provision**

**Andrea Manfrin<sup>1</sup>, Cecilio Mar Molinero<sup>2</sup>, Janet Krska<sup>3</sup>**

*<sup>1</sup>Sussex School of Pharmacy, School of Life Sciences, University of Sussex, UK, <sup>2</sup>Kent Business School, University of Kent, UK, <sup>3</sup>Medway School of Pharmacy, University of Greenwich and Kent, UK*

Asthma prevalence is increasing and the economic loss due to lack of asthma control is €72 billion in EU 28. Pharmacists have a role to play, and a bespoke novel pharmacist-led intervention for asthma patients, called Italian Medicines Use Review (I-MUR), has shown both effectiveness and cost-effectiveness. The I-MUR intervention enables asthma patients to optimise the effect of their medications. This study aimed at assessing the mismatch between patients' attitude-perception towards their medications and their complaints during the I-MUR service provision. The I-MUR was provided in four different Italian locations; data were collected and analysed using descriptive statistics, optimal scaling and contingency tables. The number of pharmacists and asthma patients involved in the study was 74 and 895 respectively. The majority of patients (72%) did not believe that they had problems with their medications, 78% confirmed that they had full knowledge and understanding of their medications, 75% said that their medications were working and 45% confirmed that they missed a dose. The number of patients who raised complaints was 683 (76%) and the number of complaints raised by each patient ranged between 1 to 5. Only 18% of the patient population reported having neither medicine-related problems nor asthma-related complaints. The use of optimal scaling and contingency tables unveiled the mismatch between patients' attitude-perception towards their medicines and the type and number of complaints raised by them during the I-MUR service provision.

**Keywords:** optimal scaling, contingency tables, asthma, patients, attitude, perception, complaints, Italian Medicines Use Review, community pharmacy

## **A Latent Class Conjoint Model for a Labour Market Analysis**

**Paolo Mariani<sup>1</sup>, Andrea Marletta<sup>2</sup>,  
Lucio Masserini<sup>3</sup>, Mariangela Zenga<sup>4</sup>**  
*University of Milano-Bicocca, Italy*

ELECTUS is a multi-centre research project with the aim to stabilize the relationship between universities and companies. Through this project, it is possible to acquire information about procedures for the entrepreneurs' choice in the recruitment process of the new graduates. Using a CAWI survey, 299 Lombardy companies with at least 15 employees were asked them to manifest their preference choosing among hypothetical profiles of new graduates with different competencies each other. The survey involves several job vacancies. The statistical methodology here adopted is Latent Class Metric Conjoint Analysis, a technique used to detect consumers' behaviour included in the more general class of Finite Mixture Models.

## **Modelling Length of Stay of Elderly Patients in Italy using Survival Methods**

**Adele H Marshall<sup>1</sup>, Ciara de Brun<sup>1</sup>, Mariangela Zenga<sup>2</sup>**

*<sup>1</sup>Mathematical Sciences Research Centre, Queen's University Belfast, Belfast, BT7 1NN, UK.*

*<sup>2</sup>Dept of Statistics and Quantitative Methods, University of Milano-Bicocca, Via B. degli Arcimboldi, 8, 20126 Milano, Italy.*

This paper considers the Coxian phase-type distribution for modelling the length of stay in hospital of elderly patients. Coxian phase-type distributions are a special type of Markov chain which describe the time which elapses until a certain event occurs as a series of sequential phases. The data consists of records from 2012 for both private and public hospitals in all 21 regions of Italy and all patients are aged 65 and over. The optimum number of phases for modelling the elderly patient data was determined to be three. It was found that the optimal three phase Coxian distribution was a better fit to the data than normal, Weibull and log-normal distributions. Survival analysis of the data was conducted using Kaplan-Meier estimates, Cox proportional hazard models and accelerated failure time models and survival trees produces. All of these approaches highlighted the covariates age, gender and urgency as significantly influencing patient survival. The survival tree also suggested patient diagnosis.

## Advanced Statistical Methods for the Analysis of Length of Studies in Higher Education Institutions

**Adele H. Marshall<sup>1</sup>, Mariangela Zenga<sup>2</sup>, Aglaia Kalamatianou<sup>3</sup>**

*<sup>1</sup>Mathematical Sciences Research Centre, Queen's University Belfast, UK, <sup>2</sup>Dept of Statistics and Quantitative Methods, University of Milano-Bicocca, Italy, <sup>3</sup>Dept of Sociology, Panteion University of Social and Political Sciences, Greece*

The purpose of this study is to highlight the effectiveness of some advanced statistical methods in the analysis of students' patterns regarding the duration of time up to a final event which may reflect on-time to graduation, late graduation or dropout by taking into account various characteristics of the students. The statistical methodology exploits a technique consisting of a sequence of stages in the survival tree using Coxian phase distributions. The results of the application of this methodology to data derived from an Italian and a Greek University converge and are quite clear and comprehensive with regard to the influence of certain characteristics of the students, especially of gender, on predicting the behavior of the students with regard to the time duration up until the final event.

## Efficient Computation of Distributions of Pattern Statistics using Probability Equivalence Classes

**Donald E.K. Martin, Laurent Noe**

*Dept of Statistics, North Carolina State University, USA*

In Markov chain based methods for computing pattern distributions, a Markov chain is set up with states indicating progress to pattern occurrences, and computation is carried out sequentially through time. After setting the initial distribution, probabilities of lying in the states at each time point are obtained by multiplying by the transition probability matrix. To compute probabilities for different parameter values, the computation is re-run. This paper forwards a more efficient method, where counts of sequences in the pattern-progress states and with constant values of sufficient statistics and the statistic of interest are updated instead of probabilities. The rationale is that the counts themselves don't depend on the parameters, and thus the set of parameters may be input into an equation after counts are obtained, saving computation time by eliminating repeated computations. The method is illustrated on the problems of computing the probability of accepting a unit in start-up demonstration tests and determining optimal spaced seeds for sequence alignment for various parameter values.

**Keywords:** Markov chain embedding, minimal deterministic finite automaton, sequence alignment, spaced seed coverage, start-up demonstration tests

## Quantile Estimation for Pareto Distribution

**Ayana Mateus, Frederico Caeiro**

*Faculdade de Ciências e Tecnologia, Universidade Nova de Lisboa, Portugal  
and Centro de Matemática e Aplicações (CMA), Portugal*

The Pareto distribution was created by the economist Vilfredo Pareto (1848-1923) who introduced it while studying distributions for modeling large incomes. Since then, the Pareto distribution has been widely applied in different fields such as finance, physics, hydrology, geology, astronomy and computer science. In this work, we compare the finite sample behaviour of several quantile estimators for a Pareto type I distribution. Quantile estimators are obtained by replacing the scale and shape parameters in the quantile expression by their estimates. For the estimation of this distribution, we shall consider several classic methods such as: the moments, the maximum likelihood and the probability weighted moments.

**Keywords:** Pareto distribution, Monte Carlo method, quantile, maximum likelihood, probability weighted moments.

### References

- [1].B.C. Arnold. Pareto and Generalized Pareto Distributions. In D. Chotikapanich (ed.) Modeling Income Distributions and Lorenz Curves, Springer, New York, 119{145, 2008
- [2].F. Caeiro and M. I. Gomes. Semi-parametric tail inference through probability weighted moments. Journal of Statistical Planning and Inference, 141, 937{950, 2011
- [3].N.L. Johnson, S. Kotz and N. Balakrishnan. Continuous Univariate Distributions, Wiley, New York, Vol.1, 2nd Ed., 1994.
- [4].R. E. Quandt. Old and new Methods of Estimation and the Pareto Distribution. Metrika, 10, 55{82, 1966
- [5].J. Hosking, J. Wallis and E. Wood. Estimation of the generalized extreme value distribution by the method of probability-weighted moments. Technometrics, 27(3), 251{261, 1985.

## Modelling Social Events through the Joint Models of Survival and Longitudinal Data

**Marcella Mazzoleni**

*Dept of Statistics and Quantitative Methods, University of Milano-Bicocca, Italy*

The joint models analyse the effect of longitudinal covariates onto the risk of an event. The sub-model compose the joint models: the longitudinal and the survival sub-models. For the longitudinal sub-model a multivariate

mixed model can be proposed. Whereas for the survival sub-model, a Cox proportional hazards model is proposed, considering jointly the influence of more than one longitudinal covariate onto the risk of the event. The estimation method use the joint likelihood formulation and it is an extension of an estimation method implemented for the joint models with univariate longitudinal sub-model. This method is based on the implementation of an Expectation-Maximisation (EM) algorithm, where in the M-step a one-step Newton-Raphson update is used, as for some parameters estimators, it is not possible to obtain closed-form expression. In addition, a Gauss-Hermite approximation is applied for some of the integrals involved. These models can be widely used in social science to analyse the probability of an event that can positively affect the life of a subject, for instance time to graduation or time for obtaining a work position. Accordingly the aim of joint models if to quantify the effect of longitudinal or repeated measures onto the probability of the event analysed. In addition, the dynamic predictions at different fixed time are shown.

**Keywords:** Joint models, Dynamic predictions, Social science.

## Improving the Computational Efficiency for Calculating Matrix Exponentials in R Using ‘Kexpmv’

**Meabh G. McCurdy, Adele H. Marshall, Karen J. Cairns**

*Mathematical Sciences Research Centre, Queen's University Belfast, UK*

The calculation of matrix exponentials play a key role in finding solutions to differential equations and are important within the field of Markov models. This is one of the main reasons why this matrix exponential operator is arguably the most widely used and widely studied matrix function. Computational complexity however, is a major issue surrounding this matrix operator which can limit its use within different application areas. This issue gets progressively worse as the dimensions of the matrix increase. The standard algorithms such as the scaling and squaring become computationally problematic when dealing with large sparse matrices. An alternative method for calculating matrix exponentials are Krylov subspace methods. These methods can be used to calculate both the matrix exponential in isolation as well as the matrix exponential with the product of a vector. The matrix-matrix operations in these methods are replaced with matrix-vector operations, which help with the computational issue. The software Expokit implements these Krylov routines to enable the efficient calculation of the exponential of both small dense matrices and large sparse matrices. The Expokit software can be implemented in both Fortran and Matlab. A new R package ‘kexpmv’ is introduced and utilised as part of this research. The aim of this package is to improve the computational efficiency for calculating matrix exponentials by

implementing the Expokit software within an R environment. An application in biogeography investigating the spread of *Vireya* *Rhododendrons* throughout Southeast Asia, along with simulated data, will be used to demonstrate the improvements made by 'kexpmv' to the computational efficiency for calculating matrix exponentials.

**Keywords:** Matrix Exponential, Krylov Subspace, Computational Efficiency

## **A Study of the Implications of the Ballast Water Management Convention on the Economy**

**Georgios I. Michalelis, Mary Karadima**  
*Athens, Greece*

The Ballast Water Management Convention entered into force on 8 September 2017, approximately 13 years after its adoption by the International Maritime Organization (IMO). The significance of the Convention is that it addresses major environmental and economic issues arising from the discharge of Ballast Water used by Ships during their operations and aims at reducing the overall impact on marine biodiversity, human health and the economy. To achieve its goals, the Convention introduces the requirement for all ships to install Ballast Water Treatment Systems onboard and use these systems when discharging their ballast water into the environment. By doing so the invasive species contained in the ballast water (e.g. the European Crab or the *Vibrio Cholera* Bacteria) are neutralized before exiting the ship, thus preventing them from affecting the local biodiversity or the human health. Ballast Water Treatment Systems, however, differ significantly from one another and their efficacy and efficiency in real time situations cannot be accurately estimated in a controlled environment testing. For example, active technologies (e.g. use of chemicals) can pose a threat to the local biodiversity themselves, if the substances used exceed certain levels, while passive technologies (e.g. UV) may not be as effective depending on the type of water, the quantity of invasive species and the water clarity. These issues can affect the economy in many ways, especially through the cost impact on Product Prices and the Capital Expenditure (CAPEX) and Operating Expenses (OPEX) requirements of shipowners. Using available information from online sources and the relevant legal framework, we examine the possible economic implications that the BWM Convention will have on the economy, using as variables the life-cycle cost, the indirect operational costs and the inspection costs related to a Ballast Water Treatment System.

**Keywords:** Economic Impact, Environmental Impact, Ballast Water Treatment.



**Acknowledgments:** *This work is part of a larger project that was prepared for a business initiative. The technical aspects are an evolution of a program funded by the General Secretariat for Research and Technology, adapted for commercial use in accordance with the IMO Convention and the US Coast Guard Regulations.*

## On a Family of Bivariate Birth Processes and Applications

**Leda D. Minkova**

*Faculty of Mathematics and Informatics, Sofia University, Bulgaria*

In this paper we introduce a family of bivariate birth processes and show that it is constructed by trivariate reduction method. As examples we define the bivariate Polya-Aeppli process (BPAP) and the bivariate noncentral Polya-Aeppli process (BNPAP). The probability mass functions, recursion formulas and some properties are derived. As application we consider a bivariate risk model with the defined counting processes. The ruin probability and the case of exponentially distributed claims are analyzed.

**Keywords:** Bivariate birth process, Polya-Aeppli process, bivariate risk model, ruin probability

## A Mixed Hidden Markov Model for Unbalanced Longitudinal Data

**Hannah Mitchell**

*Queen's University Belfast, UK*

Hidden Markov models have been utilised in many different application areas from speech recognition to fraud detection and gene therapy. They describe the relationship between two stochastic processes, whereby they make inference about an unobserved process based on an observed one. In recent years there has been a rise in popularity of longitudinal studies in fields such as sociology and medicine. This type of data arises from the repeated measurement of the same subjects over time. Hidden Markov models have been developed for use with longitudinal data as they can account for features which are inherent within this type of data such as serial dependence and heterogeneity in the subjects.

However, longitudinal data typically will be unbalanced due to many reasons. The continuous time hidden Markov model is a potential way to deal with this type of data, as it allows for the occurrence of observations at arbitrary times. This research explores the potential use of the continuous time hidden Markov model for unbalanced longitudinal data. In particular the continuous time hidden Markov model with linear mixed.

**Keywords:** Hidden Markov model, Linear mixed effects, Longitudinal model, Unbalanced data effects will be discussed.

## Damage Analysis Induced by 2 Stochastic Processes

Hiroaki Mohri<sup>1</sup>, Jun'ichi Takeshita<sup>2</sup>

<sup>1</sup>Faculty of Commerce, Waseda University, Japan, <sup>2</sup>Research Institute of Science for Safety and Sustainability, National Institute of Advanced Industrial Science and Technology (AIST), Japan

In this study, we investigated in a damage induced by 2 types of stochastic process against one unit (later we shall treat units and networks e.g. water pipeline or electricity networks).

We suppose the 2 types of damages.

s1) continuous type: Failures may occur, as the use of age. (e.g. Corrosion of metal pipe progress)

s2) shock type: Failures may occur at unexpected events like earthquakes.

On an unit (e.g. metal pipe), we suppose the 1st stochastic process as its life period. It would be a stochastic process of "continuous type". And the 2nd stochastic process is counting process as shock events.

Firstly, each unit has 1 point initially. If the point decrease and reach 0 as s1 type damage progress, its life period would end. If it was shocked by s2 type damage, the point would be 0 simultaneously even if within the life period.

For application point of views, its life time may follow Weibull distribution. Also, s2 is Poisson process. Period between shock events may follow exponential distribution.

Even tough, using Weibull distribution is hard to analyze for this problem. We use Gamma (Erlang) distribution approximately instead of Weibull distribution.

Damage analysis here if time  $t$  is given, we may show probability of damage. Also, we may calculate the first passage time to failure.

Now, not only 2 stochastic process (i.e. Poisson Process or Gamma(Erlang) process) but also we consider shock weight for each event. (we will define shock for s1 "continuous type" in our paper).

$W$ : the amount of Shock for s1 and its distribution function  $G_W(x)$ .

$U$ : the amount of Shock for s2 and its distribution function  $H_U(x)$ .

Therefore, we treat compound Poisson Process and compound Gamma Process for this problem.

In this talk, we shall have damage analysis of the 2 compounded stochastic process. In some easy case, we will show the optimal replacement policy by applying the reliability and maintenance theories by Barlow[1] and Nakagawa[2][3] for this problem.

**Keywords:** Reliability, Applied Stochastic Process, Networks

**References:**

- [1].R. E. Barlow and F. Proschan, Mathematical Theory of Reliability, Wiley, New York (1965).
- [2].T. Nakagawa, Random Maintenance Policies, Springer, London (2014).
- [3].T. Nakagawa, Shock and Damage Models in Reliability Theory, Springer, London (2007).
- [4].J. Takeshita and H. Mohri, Investigation of a Network Failure Problem with a Significant Path, from the Perspective of Crisis Management, Journal of Mathematics and System Science 4 (2014) 479-485.

## The Generalized Vandermonde Interpolation Polynomial Based on Divided Differences

**Asaph Keikara Muhumuza<sup>1,2</sup>, Karl Lundengård<sup>1</sup>, Jonas Österberg<sup>1</sup>,  
Sergei Silvestrov<sup>1</sup>, John Mango<sup>3</sup>, Godwin Kakuba<sup>3</sup>**

<sup>1</sup>*Division of Applied Mathematics, Mälardalen University, Sweden,* <sup>2</sup>*Dept of Mathematics, Busitema University, Uganda,* <sup>3</sup>*Dept of Mathematics, College of Natural Sciences, Makerere University, Uganda*

In this article, we will construct the divided differences interpolation polynomial based on the generalized Vandermonde determinant approach. Some results regarding the appropriateness for this method for curve-fitting and approximation will be discussed. The proposed interpolation technique will be tested by construction of approximative models based on experimental data.

**Keywords:** Generalized Vandermonde determinant, Divided Differences interpolating polynomial, Approximative models.

## Malnutrition>Anemia>High MMR in India: Some Analysis

**Barun Kumar Mukhopadhyay**

*Population Studies Unit, Indian Statistical Institute, India*

The Global Nutrition Report 2017 (the first report published in the month of November, 2017) presented at Milan, Italy clearly gives a grim nutritional status of Indian mothers with 51 per cent suffering from anemia and 22 percent overweight in addition to many other countries' situation on the global aspect. Almost similar observations were made before five years back in 2012 by the International Food Policy Research Institute (IFPRI), USA while prescribing their 1,000 day window about India. Keeping in view of these reports and suggestions Indian government no doubt adopted some schemes and policies. Unfortunately there are at the same time some lacuna in India with huge population of almost 1.33 billion along with the largest democracy in the 00world and diversities in many aspects including geographical regions, caste, creed, religion and

especially mother tongue. In the past some studies have been made by different authors (Mukhopadhyay, 2015; Garg et al, 2010) relating to high Maternal Mortality. The former observed two causes, high rate of domestic delivery and early pregnancy in Indian data while the former also observed similar findings about home delivery in some major state of India. In the present paper an attempt is made to find whether low calorie, carbohydrate based diet with anemic mothers contribute to high MMR in India apart from other two causes as mentioned above. The analysis are under process and whatever final findings will come no doubt may give impetus to GOI in order to move at least one step further in achieving the target of Sustainable Development Goal (SDG) by 2030. The data will be used from sources including Indian Census, SRS and different net surfing.

**Keywords:** MMR, SDG, Census, SRS and net surfing

### **Variable-Geometry Families in Italy: A Statistical Description**

**Ilia Negri, Gisella Bassanini**  
*University of Bergamo, Italy*

In the last 40 years the structure of the Italian family has changed. It is an important issue to find the data to describe numerically as this change has taken form. There are difficulties also in find the correct definition of what a family is when the form of the family is not the one expected from the common sense.

In particular in Italy a family where there is only the mother and one child without a father are not census as a category but are census as a never married woman with a child. In this last group anyway also mother with child that separate with the father never married are counted.

Have an exact imagine of how many different family we have is important for institutions to start with policy directed to these particularly fragile (economically and emotionally) category of families.

### **Social Application with Multivariate Regression Chain Graph Models**

**Federica Nicolussi, Manuela Cazzaro**  
*Dept of Statistics and Quantitative Methods, University of Milano Bicocca, Italy*

This work studies how the different level of education can be affected by individual characteristics and, at the same time, how the scholastic training can affect other social aspects of the interviewees' life. We propose to describe this kind of relationships through a multivariate logistic

regression model based on the Chain Graph model. By following the approach of Marchetti and Lupporelli, [2], in fact, we take advantage from a particular case of Chain Graph model, called ‘of type IV’, in order to express variables as *purely explicative*, *purely response* and *mixed variables*. In addition, we also study the relationships under the context-specific independence point of view. This means that we study if there are conditional independencies that hold only for a subset of modalities of the conditioning variables. Formally, a context-specific independence has the form  $A \perp B|C = i_c$  where  $A$ ,  $B$  and  $C$  are three sets of variables and  $i_c$  is the vector of certain modalities of the variables in  $C$ . Nyman et al. [4] handle with the context-specific independencies in the graphical models, through the so-called *strata* added to the graphs. We improved their approach by implementing the strata also in the Chain Graph models, see [3]. This work is finalized in showing the multiple aspects that it is possible to highlight by implementing these models, in both graphical and parametric point of views. The analysis were carried out on the ISTAT dataset on the ‘*aspects of everyday life*’, [1].

**Keywords:** Chain Regression Model, Multivariate Logistic Regression Model, Context-specific independence, Educational Study.

#### References

- [1]. Istat. Multiscopo ISTAT - *Aspetti della vita quotidiana*. UniData - Bicocca Data Archive, Milano. Codice indagine SN167. Versione del file di dati 1.0 (2015)
- [2]. Marchetti, Giovanni M., & Lupporelli, Monia. *Chain graph models of multivariate regression type for categorical data*. Bernoulli, 17(3), 827-844, (2011).
- [3]. Nicolussi, Federica & Cazzaro, Manuela. *Context-specific independencies for ordinal variables in chain regression models*. arXiv:1712.05229, (2017).
- [4]. Nyman, Henrik, Pensar, Johan, Koski, Timo, & Corander, Jukka. *Context specific independence in graphical log-linear models*. Computational Statistics, 31(4), 1493-1512, 2016.

## Investment Timing when Investment Opportunities Arrive in a Random Sequence

**Michi Nishihara**

*Graduate School of Economics, Osaka University, Japan*

This study considers a firm's optimal investment timing problem when investment opportunities arrive in a random sequence and are irreversible. We analytically derive the project value and the investment threshold. Our results converge to those of the real option value (ROV) method as the arrival rate of investment opportunities is higher, whereas our results converge to those of the net present value (NPV) method as the arrival rate is lower. Further, we extend the results to a case with two types of investment opportunities, namely low- and high-growth opportunities. We analytically derive the condition under which the firm always forgoes low-growth opportunities. A notable result is that the firm accepts a lower-

growth opportunity for a low arrival rate and a high state variable. Our results build a bridge between the NPV and ROV methods.

**Keywords:** Real option; Net present value; Illiquidity.

## Educational Data Mining and Learning Analytics: state of the art and challenges

Amílcar Oliveira<sup>1,2</sup> and Teresa Oliveira<sup>1,2</sup>

<sup>1</sup>Dept of Sciences and Technology, Universidade Aberta, Lisbon, Portugal

<sup>2</sup>CEAUL - Center of Statistics and Applications, University of Lisbon, Portugal

Education is one of the most important and most discussed issues in all of society. Being part of a process of acquiring knowledge and / or skills, it has undergone a great number of changes over time.

In particular, in the last decade, advances in information and computer technology have allowed greater interaction between the participants in the teaching / learning process.

All recent technological developments have made it possible for both school and university stakeholders to change their way of transmitting and sharing knowledge. In many cases, particularly at the level of higher education, the changes go as far as the methodology used in teaching their courses where, in addition to traditional teaching, other teaching modalities are adopted: b-learning (teaching with presence and distance components) and / or e-learning (fully online and distance learning).

Teaching / learning models based on online learning environments allow students to gain access to knowledge anytime, anywhere, which appear to be significant advantages. In opposition to these advantages, this modality is faced with challenges, namely due to the absence of synchronous communication, direct human contact and the gaps that this reality can generate. Still, e-learning advocates argue that the creation of virtual communities that interact in different forms: chats, forums, messages through e-mail, among others, compensate this lack, enriching the relational process between individuals with the same interest, yet with different views and located in different regions and / or countries.

With the significant increase in the use of online teaching and learning environments and other supporting technologies, large volumes of data are generated and stored in different areas of the systems used, involving in particular students and teachers. Through the monitoring and analysis of these data, it is possible to obtain important information, both in quantity and quality, with a view to improving the quality of teaching, as well as combating failure and dropping out of school.

This presentation will focus on possible ways of analysing data based on statistical techniques, in order to monitor, extract relevant knowledge and predict school performance in online teaching.

**Keywords:** Education, Data Mining, Learning Analytics, Statistical Techniques.

### References

- [1]. Baker, R. (2010) Data Mining for Education. Em: McGaw, B., Peterson, P., Baker, E. (eds.), International Encyclopedia of Education (3rd edition), vol. 7. Oxford, UK: Elsevier. 112-118.
- [2]. G. Siemens, R.S.j.d. Baker (2012). "Learning analytics and educational data mining: towards communication and collaboration". Proceedings of the 2nd International Conference on Learning Analytics and Knowledge: 252–254.
- [3]. Juan, A., Minguillón, J., Huertas, A., Sancho, T. (2011) Computer-supported statistics courses in online environments: adding e-repositories to the equation. Em: Int. J. Teaching and Case Studies, 3: 16-34.
- [4]. Miguel Á. Conde; \_Angel Hernández-García; Amílcar Oliveira (2015) Endless horizons? Addressing current concerns about learning analytics. Proceedings of TEEM15-Third International Conference on Technological Ecosystems for Enhancing Multiculturality, 259-262, Gustavo R. Alves and Manuel C. Felgueiras (Eds.)
- [5]. R Development Core Team (2016) R: A Language and Environment for Statistical Computing. The R Foundation for Statistical Computing, Vienna, Austria. URL:<http://www.r-project.org/>.
- [6]. Romero, C., Ventura, S. (2013) Data Mining in Education. Data Mining Know Discov, 3: 12-27
- [7]. Romero, C., Ventura, S. (2010). Educational Data Mining: A Review of the State-of-the-Art. **IEEE Transaction on Systems, Man, and Cybernetics, Part C: Applications and Reviews.** 40(6), 601-618.

## Some Combinatorial Designs Using R

**Amílcar Oliveira<sup>1,2</sup> and Teresa Oliveira<sup>1,2</sup>**

<sup>1</sup>*Dept of Sciences and Technology, Universidade Aberta, Lisbon, Portugal*

<sup>2</sup>*CEAUL - Center of Statistics and Applications, University of Lisbon, Portugal*

The mathematical issues on Experimental Design were fostered and developed by Fisher and his colleagues in the 1920s, giving rise to combinatorial design theory, a field intimately linked to its applications. The Combinatorial Design Theory is crucial on the search for the allocation of the elements of a finite set into smaller subsets, in order to ensure that certain balance properties are satisfied. Powerful combinatorial and computational techniques for the construction of a huge diversity of models of Experimental Designs obeying such balance properties are required.

Balanced Incomplete Block Designs (BIBD) are very well known not only by the balance characteristics but also thanks to their optimal properties. Indeed this field grew so far and so fast that its historical connection to applications was strained.

Many real problems revealed that it is not easy to overcome the difficulties in constructing BIBD with certain combinations of parameters. Thus it is important to consider and to analyze other designs somehow related to BIBDs but with relaxation of some imposed restrictions. Also it is important to investigate the development created in the R software for analysis of

such models in order to attain for eventual new contributions. In this work BIBD relations with other types of designs are explored - Latin Squares Designs and Partially BIBD. Illustrations of different applications in several areas are presented using R.

**Keywords:** Combinatorial Design, Experimental Design, BIBD, Latin Square, PBIBD, software R.

#### References

- [1]. Charlie Colbourn and Jeff Dinitz (2006). Handbook of Combinatorial Designs, Second Edition. Chapman & Hall, ISBN: 1-58488-506-8.
- [2]. Oliveira, T.A. (2010). BIBD with Repeated Blocks: Review and perspectives. Proceedings of ICCS-X Conference, Tenth Islamic Countries Conference on Statistical Sciences-Statistics for Development and Good Governance. Editors Zeinab Amin and Ali, S., Hadi, The American University in Cairo. ISBN 978-977-416-365-8. Volume I, 82-96.
- [3]. R Development Core Team. R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria, ISBN 3-900051-07-0 (2012).

### Evaluating the Role of Factors' Rotation in Multivariate Factor Analysis, Regarding the Repeatability of the Information Derived, under Various Scenarios of Random Error

**Dimitris Panaretos, Malvina Vamvakari, George Tzavelas,  
Demosthenes Panagiotakos**

*School of Health Science and Education, Harokopio University, Greece*

In the past years, the need for processing large amount of data in health science, and particularly in molecular biology, genetics and epidemiology (e.g., nutrition patterns, successful aging patterns), has arisen. This created the need to apply advanced analytical methods that could enable researchers to handle such amount of information. Principal Component Analysis (PCA) and Factor Analysis (FA) are largely used as dimension reducing procedures. A technique that may help to better retrieve "valuable" information from these multivariate methods is the rotation of the axes. It has been suggested that rotation of the axes is required so that the extracting components/factors can be more "interpretable" due to the fact that rotation maximizes the variance explained of the extracted components and makes the pattern of loadings more well-defined. The rotation can be orthogonal (the factors are uncorrelated) or non-orthogonal (the factors are correlated). The common methods of orthogonal rotation are Varimax, Quartimax and Equimax. Even though these techniques have been used widely, it seems to be a large gap in the literature regarding the repeatability of the extracted factors, when random error exists. The aim of this paper is to study the role rotation, as regards the repeatability of the extracted factors under various scenarios of



random error in the initial information. In particular, a set of data ( $x$  variables \*  $n$  cases) will be simulated from the Normal or the Uniform distribution, and random error will be imputed in each variable at a second stage, in an attempt to reflect a “real world” case scenario. In both stages, FA with and without rotation will be applied and the repeatability of the extracted factors will be tested in each case and under various rotation methods. This procedure will be repeated 1000 times and the results of the repeatability tests will be summarized under the various rotation methods.

**Keywords:** factor analysis; multivariate analysis; rotation; repeatability

## **Subjective Survival: An Assessment of Accuracy according to Individual Social Profile**

**Apostolos Papachristos, Fia Cera, Georgia Verropoulou**

*Dept of Statistics and Insurance Science, University of Piraeus, Greece*

Past literature suggests that personal views on subjective survival vary according to an individual's characteristics. The present study aims at identifying traits that relate to the underestimation or the overestimation of survival, using information from an especially rich dataset, the Survey of Health Ageing and Retirement in Europe (SHARE wave 6, 2015) and the Human Mortality Database. The statistical analysis is based on generalized linear regression models.

The findings indicate that better cognitive function and higher educational attainment are linked to more accurate predictions. On the other hand, female gender, smoking, poor health and memory, lower income and higher depression levels seem associated with the under-estimation of future life expectancy. Further, persons who consume less frequently fruits or vegetables and eggs or legumes, also tend to under-estimate future life expectancy. By contrast, older age, better quality of life, being in paid employment and having more children are related to the over-estimation of survival. Furthermore, widowed, divorced, never married and financially distressed individuals also tend to over-estimate future life expectancy.

Hence, the study indicates, in line with past research, that accuracy of predictions diversifies across individuals with different characteristics. Grouping of traits can be of particular interest for life insurers and pension funds to be incorporated in their benefits' structure and policies.

**Acknowledgement:** *This work has been partly supported by the University of Piraeus Research Center.*

## State Occupancies, First Passage Times and Duration in DNA sequences via semi Markov modelling

**Aleka Papadopoulou, Haris Palikrousis and Evanthia Farmakioti**

*Dept of Mathematics, Aristotle University of Thessaloniki,  
Thessaloniki 54124, Greece*

The aim of this paper is to provide analytic relations for three types of probabilities that are of considerable research interest in semi Markov modelling. First, we study the probabilities of the state occupancies for the non homogeneous case. For this purpose, we use the basic recursive relation to provide a description of the corresponding geometric transforms and then by applying related properties we produce relations for the probabilities in closed analytic form i.e. in relation to the basic parameters of the model. Furthermore, a verification of the results is performed numerically and by comparison of the corresponding types which are obtained by the analytic and recursive relations. Finally, an illustration is realized in Mathematica V.9 by applying data from a DNA sequence. Also, we present two more important types of probabilities: Probabilities of the first passage time and the duration probabilities. We define their corresponding recursive equations for the non homogeneous case and provide all the theorems with which we can find the closed analytic forms for the above probabilities. Last, an application is realized in Matlab (V. R2015a) to illustrate the previous theoretical results by applying data from a DNA sequence.

**Key Words:** Semi Markov Chains, State Occupancies, First Passage Times, Duration, DNA Sequences.

## Sampling in Special Kinds of Nets

**Papageorgiou Myrto, Farmakis Nikolaos**

*School of Mathematics, Aristotle University of Thessaloniki*

In many cases of sampling applications the (target) Population has the form of a net. In this case we face some problems especially with the bias of estimators, as we use classic sampling procedures, like simple random sampling (srs), systematic sampling, etc. In the present paper we try srs in nets, providing unbiased estimators of the mean value of some random variables. This result is reached by adoption of some axioms for the nets. A set of five axioms is associated with the net. These suitable nets are taken to have the form of the graphos presented by a matrix or, at least, by a part of the mentioned matrix. A theoretic proof for the unbiasedness of the estimators is given in a graphos with size the nature number:  $N=4,5,\dots$ . The graphos constituted by the above axioms works as a kind of finite geometry. We could define some kind of parallelism in a structure like this

geometry. Some examples are given to illustrate the above ideas of sampling in nets with the structure of finite geometry and show that the estimators of the parameters (mean value at least) are unbiased.

**Key Words:** Net, graphos, sampling, unbiased,  
MCS2010: 62D05, 97K20, 05A19, 05B20,

#### References

- [1].Artemiadis N. (1977) "Elementary Geometry from an upper point of view", Publications of Hellenic Mathematical Society, Athens (in Greek).
- [2].Farmakis N. (2017) "Introduction to Sampling", Kyriakidis Bros Editing S.A, Thessaloniki (in Greek).
- [3].Farmakis N. (2015) "Sampling and Applications" Hellenic Academic e-Books, Athens, (in Greek). ISBN: 978-960-603-093-2.
- [4].CochranW. (1977) "Sampling Techniques", John Wiley & Sons, Inc, New York, Toronto.

## A study on Probability Density Functions with Negative Exponent

**Papatsouma Ioanna, Farmakis Nikolaos**

*Dept of Mathematics, Aristotle University of Thessaloniki, Greece*

Sampling can help researchers to study the distributions of the various random variables and produce very satisfying results in terms of accuracy and speed. The probability density function (pdf) of a random variable (rv)  $X$  can be estimated via sampling and various parameters of the rv  $X$ , e.g. the mean, the variance, the coefficient of variation, the range, etc. and then with knowledge of the pdf, further statistical problems can be tackled. This procedure is both theoretical and empirical, based on sample data. In applied statistics, an initial hypothesis is made upon for a particular situation about the form of the pdf of  $X$  or an approximation of this pdf, that can be efficient and easy to be managed by researchers. The choice of the type of the suitable pdf and the time it takes to process the data in order to construct the pdf is vital. In the present paper we study how to get the suitable estimator from the sample, when the pdf has a polynomial form with a negative exponent. Illustrative examples are given to highlight the basic results of the theoretical approach.

**Keywords:** Sampling, Probability density function, Coefficient of variation, Mean, Variance, Range.

**MSC2010 Classification:** 62D05, 62E17

## Periodic Regression Models for Outbreak Detection in Biosurveillance Systems

Christina Parpoula<sup>1</sup>, Alex Karagrigoriou<sup>1</sup>, Angeliki Lambrou<sup>2</sup>

<sup>1</sup>Lab of Statistics and Data Analysis, Dept of Mathematics, University of the Aegean, Greece, <sup>2</sup>Hellenic Center for Disease Control and Prevention, Greece

In recent years, the emergence of new infectious diseases and the occasional rapid increase of their cases worldwide, the intense concern about bioterrorism, pandemic influenza or/and other Public Health threats, and the increasing volume of epidemiological data, are all key factors that make necessary the development of advanced biosurveillance systems. Additionally, these factors have resulted in the awakening of the scientific community for introducing new and more efficient epidemic outbreak detection methods. A rich array of statistical methodologies is available for the early detection of epidemic activity in biosurveillance systems. This alone raises the research question “Which is the most appropriate methodology to use?” It is not feasible to make detailed recommendations as to which method is “best”, because this depends critically on the specific details of the application and implementation, as well as on its purpose and context. In this work, baseline periodic regression models are fitted for establishing “normal behavior”, and then based on an “anomaly detection scheme”, epidemic alert thresholds are produced which can be used to perform real time surveillance. Following a model selection pathway, the periodic regression model chosen as the optimal one for a study of weekly influenza-like syndrome cases succeeded in detecting the beginning and end of the epidemic waves, identified the pattern that best matches the recent influenza activity, and outperformed typical forecasting models. Improved baseline modelling, and subsequent reduction of residual errors, led to improved outbreak detection.

**Keywords:** statistical modelling, data analysis, periodic regression model, computer science, biosurveillance, epidemiology.

### References

- [1].C. Pelat, P.-Y. Boëlle, B.J. Cowling, F. Carrat, A. Flahault, S. Ansart and A.-J. Valleron. Online detection and quantification of epidemics. BMC Medical Informatics and Decision Making, 7:29, 2007.
- [2].G. Shmueli and H. Burkom, H. Statistical challenges facing early outbreak detection in biosurveillance. Technometrics, 52, 39–51, 2010.

## Quantile Hedging for Annuity Contracts

**Jevtic Petar<sup>1</sup>, Traian Pirvu<sup>2</sup>, Tuan Q. Tran<sup>3</sup>**

<sup>1</sup>Arizona State University, USA, <sup>2</sup>McMaster University, Canada, <sup>3</sup>Ryerson University, Canada

Within the context of the Solvency II directive, annuity portfolio managers are charged with the obligation to maintain a portfolio composition that ensures, for a given time horizon, the solvency of the fund with stated probability. Often, in practice, given the stochastic nature of all asset classes as well as mortality, it is not clear how this can be achieved. From a practical perspective, this is a fundamental problem. In this paper, in the simple setting of one asset and purely annuity portfolio whose aggregate liabilities evolve according to stochastic survivorship of the annuitants, we investigate the dynamic optimal investment strategy between the riskless and the risky asset such that the portfolio remains solvent with stated probability before the extinction of the fund. We model the return of the risky asset with Geometric Brownian Motion and, in contrast to existing literature, we accommodate mortality evolution in continuous framework with mortality intensity following affine processes. Representative numerical examples are provided.

## Longevity Bond Pricing in Equilibrium

**Traian Pirvu<sup>1</sup>, Jevtic Petar<sup>2</sup>, Minsuk Kwak<sup>3</sup>**

<sup>1</sup>McMaster University, Canada, <sup>2</sup>Arizona State University, USA, <sup>3</sup>Hankuk University of Foreign Studies, Korea

We consider a partial equilibrium model for pricing a longevity linked bond in a model with stochastic mortality intensity that affects the income of economic agents. The agents trade in a risky financial security and in the longevity linked bond in order to maximize their utilities. Agent's risk preferences are of monetary type and are described by BSDEs (backward stochastic differential equations). The endogenous equilibrium bond price is characterized by a BSDE. By using Clark-Haussmann formula, we prove that the longevity bond completes the market.

## Information Measure for Financial Time Series: Quantifying Short-Term Market Heterogeneity

Linda Ponta<sup>1</sup> and Anna Carbone<sup>2</sup>

<sup>1</sup>LIUC - Cattaneo University, Castellanza, Italy

<sup>2</sup>Politecnico di Torino, Torino, Italy

A well-interpretable measure of information has been recently proposed based on a partition obtained by intersecting a random sequence with its moving average. The partition yields disjoint sets of the sequence, which are then ranked according to their size to form a probability distribution function and finally fed in the expression of the Shannon entropy. In this work, such entropy measure is implemented on the time series of prices and volatilities of six financial markets. The analysis has been performed, on tick-by-tick data sampled every minute for six years of data from 1999 to 2004, for a broad range of moving average windows and volatility horizons. The study shows that the entropy of the volatility series depends on the individual market, while the entropy of the price series is practically invariant for the six markets. Finally, a cumulative information measure - the  $\epsilon$ m Market Heterogeneity Index - derived from the integral of the entropy measure, is introduced for obtaining the weights of an Efficient Portfolio. A comparison with the weights obtained by using the Sharpe ratio - a traditional risk diversity measure - is also reported.

## The Health Policy and the Recent Trends in Mortality in Poland

Potrykowska Alina

*The Government Population Council of Poland, Poland*

The health policy should, contribute to the implementation of the population objectives in Poland, as it has reached a critical stage of its demographic development. For a couple of years, the number of residents has not been growing, which results from the generally zero, and in some years even negative, balance between the number of births and deaths. At the same time, for almost the entire post-war period, Poland has been losing its population for the benefit of other countries, especially Western European and the USA.

These unfavourable demographic development features regarding births and deaths are also largely related to the level and quality of public health. The most significant disadvantages displayed by Poland in various areas of public health, as compared to Western European countries are: the average life expectancy of women in Poland is still by 8.5 years longer than the average life expectancy of men, and this gap is much bigger than in the countries with the lowest mortality rate. Despite the low level of

deaths among newborn children, it is still twice as high as in the European countries which are leading in this respect.

The level of premature deaths, i.e. before the age of 65 years, is still by 50% higher than in the EU countries. This is related to the level of mortality due to circular diseases, cancer or the so-called extraordinary deaths, which in Poland is still higher than in the reference countries.

This paper presents an analysis of the cancer incidence and mortality in Poland.

## **Resampling Procedures for a More Reliable Extremal Index Estimation**

**Dora Prata Gomes, Manuela Neves**

*FCT, and CMA, Universidade Nova de Lisboa, Portugal*

Extreme Value Theory (EVT) deals essentially with the estimation of parameters of extreme or rare events. A large number of applications in areas such as biology, environment, finance, hydrology and telecommunications reveals the importance of adequate estimation procedures.

The key parameters in EVT are the extreme value index (EVI) and the extremal index (EI). Under a framework related to large values, the EVI measures the right tail-weight of the underlying distribution and The EI characterizes the degree of local dependence in the extremes of a stationary sequence. This parameter needs to be adequately estimated, not only by itself but because its influence on other parameters, such as, a high quantile, the return period, the expected shortfall. Like other semi-parametric estimators, EI estimators show nice asymptotic properties, but a high variance for small values of  $k$ , the number of upper order statistics used in the estimation, and a high bias for large values of  $k$ . This brings a real need for the choice of  $k$ .

After a brief reference to some estimators of the EI and their asymptotic properties, we mention the importance of resampling techniques in deriving reliable semi-parametric estimators. Block-bootstrap and Jackknife-After-Bootstrap are two computational procedures applied here for improving the behaviour of the extremal index estimators. An adaptive choice algorithm for the block size for the resampling procedure as well as for the choice of the more adequate number of upper order statistics for the estimation is studied. Results from an intensive simulation study are shown. Applications of these procedures to the analysis of environmental and financial data are undertaken.

**Keywords:** Extremes, Extremal Index Estimation, Resampling Procedures, Analysis of Environmental Data and financial data

**Acknowledgment:** This research was partially supported by National Funds through FCT - Fundação para a Ciência e a Tecnologia, Portugal, project UID/MAT/00006/2013 (CEAUL) and PEst-OE/MAT/UI0297/2013 (CMA/UNL).

## One-Dimensional Discrete Scan Statistics Associated to Some Dependent Models

**Cristian Preda, Alexandru Amarioarei**

*Dept of Mathematics, University of Lille and Inria Lille/Modal, France*

We consider the one-dimensional discrete scan statistic generated by a block-factor type model. Under the null hypothesis, in which the block-factor model is obtained from an i.i.d. sequence of random variables, we present an approximation for the distribution of the scan statistics and the corresponding error bounds. Numerical results are presented to evaluate the accuracy of the approximation.

**Keywords:** scan statistics, m-dependance, block-factor

## Risk-Adjusted Control Charts: A Review

**Stelios Psarakis<sup>1</sup>, Athanasios Sachlas<sup>1,2</sup>, Sotiris Bersimis<sup>2</sup>**

*<sup>1</sup>Dept of Statistics, Athens University of Business and Economics, Greece,*

*<sup>2</sup>Dept of Statistics and Insurance Science, University of Piraeus, Greece*

In the last two decades a modification of standard and advanced control charts appeared in the bibliography to improve the monitoring mainly of medical processes. This is the risk-adjusted control charts which take into consideration the varying health conditions of the patients. In this work we have tried to present all the risk-adjusted control charts presented in the literature. The risk-adjusted charts have been grouped into four categories: control charts for continuous variables, control charts for attributes, time-weighted control charts, and multivariate control charts. The application of risk-adjusted control charts in practical medical processes is also discussed.

### References

- [1]. P. Biswas and J.D. Kalbfleisch. A risk-adjusted cusum in continuous time based on the cox model. *Statistics in Medicine*, 27:3382-3406, 2008.
- [2]. J.C. Benneyan and A.D. Borgman. Risk-adjusted sequential probability ratio tests and longitudinal surveillance methods. *International Journal for Quality in Health Care*, 15(1):5-6, 2003.
- [3]. L.H. Sego, M.R. Reynolds Jr, and W.H. Woodall. Risk-adjusted monitoring of survival times. *Statistics in Medicine*, 28:1386-1401, 2009.



## **Modeling the Reliability and Performance of a Wind Farm Using the Universal Generating Function Technique**

**Panagiotis Psomas, Agapios N. Platis, Vasilis P. Koutras**

*Dept of Financial and Management Engineering, School of Engineering,  
University of the Aegean, Greece*

The objective of this paper is to assess the dependability of a wind farm power system and particularly the overall ability of the system to generate and supply electrical energy using the Universal Generating Function Technique (UGF). There are two major factors affecting the general performance and thus the wind farm output: the wind intensity and the wind turbine failures. For the wind intensity, a power curve is used to determine the monthly energy output for a given wind turbine in a given location. However, the energy output depends also on the different wind turbine degradation states due to various failures. Combining the wind intensity categories with the power output of a wind farm consisting of 20 wind turbines, we have developed a multi-state system model, characterizing all the different levels of energy output through a formally composition operator, to obtain the final system dependability measures, expressed by the availability and the Expected Energy Not Supplied (EENS). The data of the presented case study were provided by the Hellenic National Meteorological Service (HNMS) and they are used for evaluating the corresponding dependability indicators taking into account two different types of repair rates.

**Keywords:** Wind Energy, Reliability, Availability, Dependability, Weibull Distribution, Homogeneous Markov Chains, Universal Generating Function (UGF)

## **Indicators of Longevity in Humans**

**Robine Jean-Marie, Cubaynes Sarah**

*MMDN, INSERM U1198, EPHE/PSL & University of Montpellier, France*

Since the beginning of the 20th century, life expectancy (LE) at birth has been established as the main indicator, if not the only indicator, of human longevity. The age when 90% of the individuals are dead, used by ecologists to measure longevity in the wild, has been hardly used for the human species. Recently the modal age at death (M), indicating the most frequent or the typical age at death has been proposed as a better indicator of the human longevity, including about its change over time. To assess the limits of the human longevity, several specific indicators have been proposed to approach the “Maximum Life Span” (MLS), including the

Maximum Reported Age at Death (MRAD) and the Highest Reported Ages at Death (HRAD). Using French data, known both for their historical depth and their high quality, we compare all these indicators. Do they provide the same information about the change in the human longevity over the last two centuries?

## Efficacy of Retention Strategies for Employee Turnover

**Evy Rombaut, Marie-Anne Guerry**

*Dept Business Technology and Operations, Vrije Universiteit Belgium*

**Rationale.** The main goal of employee retention is to avoid dysfunctional turnover. When analysing the main reasons why employees leave and when determining the turnover probability, the question rises how turnover can be minimized and which retention strategies have an actual effect on turnover.

**Methodology.** To determine the efficacy of different retention strategies, an overview is given of the retention strategies that can be found in literature. Next, uplift techniques are used to test the efficacy of the different strategies. The uplift model is based on random forest estimation. In addition, subgroup analysis is conducted to be able to customize the retention strategies.

**Findings.** Almost all retention strategies are found not effective for the entire population. However, for each strategy, subgroups of the population were determined to learn which strategy works for which type of employee.

**Practical implications.** The results yield useful information for Human Resources practitioners. The subgroup analysis results in detailed retention information for these practitioners, which allows them to target each employee with the strategies that are most likely to succeed in retaining their employment.

**Originality.** With the uplift techniques, the actual effectiveness of retention strategies is tested on existing data in a Human Resources dataset without supplementary data collection.

**Keywords:** Turnover, retention strategies, uplift, subgroup analysis.

## **Assessing Age-Dependent Patterns of Mammalian Demographic Rates in the Wild using the MALDDABA Database**

**Ronget Victor, Lemaitre Jean-François, Gaillard Jean-Michel**  
*Université Lyon 1; CNRS, France*

In most mammalian species studied to date, both mortality and reproductive rates vary with age, with a pattern of variation that changes in relation to the level of analysis (i.e. individual, population, or species). Despite the increasing number of research programs aiming to estimate demographic parameters of mammalian populations in the wild, only a few studies have investigated the magnitude of the variation in age-specific demographics rates across a large range of species. One of the main reasons for this absence of analysis is the difficulty to gather the life-tables because they are scattered in the literature. Moreover, differences in methodology and quality between the published datasets make the comparison among populations or species especially challenging. To fill this knowledge gap, we gathered studies displaying age-dependent mortality and reproduction rates from wild populations of mammals. These data constitute the core of the MALDDABA database (i.e. MAMMALIAN Demographic DATAbase) that aims to be the main source of information for age-specific demographic rates and associated life-history traits in mammals so far available, which currently includes 160 mammalian species distributed all around the world.

We fitted a series of models involving different shapes of age dependence in relation to data quality (e.g. longitudinal vs. cross-sectional) and to the methods used for estimating demographic rates. The most challenging point was to devise standard metrics to assess age dependence across a large range of populations and species to be compared because the shape of the mortality curves strongly differed among populations. We propose a set of metrics to get a standard assessment of age-dependence in demographic rates of each population, including the age when 90% of the individuals are dead to measure longevity or the Gompertz rate to measure the senescence rate. Overall, our analysis reveals a wide diversity of age-specific patterns across mammalian species in both survival and reproductive rates. However, we stress the need of developing more standardized metrics that would allow us to compare age-specific survival and reproductive rates across species and populations.

## Health Status of Elderly People in City and Countryside in Russia

**Tamara Sabgayda, Anna Edeleva and Victorya Semyonova**

*Federal Research Institute for Health Organization and Informatics of Ministry of  
Health of Russian Federation*

In the Russian Federation, morbidity statistics are formed for general group of pensioners (60 years and older for men and 55 and older for women) in case of their coming to physicians. With a high proportion of ill-defined causes in the death causes structure, the state of health of persons who have reached the old age remains partly uncertain. In order to implement programs aimed at preserving the health of elderly, it is necessary to have more data on their health status. It should be understood, however, how the differences in access to health care are affected by the health of the elderly.

There was analyzed the spread of diseases among people aged 75 years and older based on results of entirely survey of all pensioners of one city area (7,808 people) and two rural areas (14,749) of the Nizhny Novgorod region. The greatest ages of men and women were 114 and 106 years in city and 99 and 116 years in countryside respectively.

Proportions of elderly men among all pensioners in urban and rural areas are close (28.1% and 27.2%) while proportion of rural women is more than urban ones (47.2% vs. 30.7%;  $\chi^2 = 394.7$ ;  $p < 0.0001$ ). Proportion of those who do not serve themselves was 3.7% among the townspeople and 3.5% among the villagers. Among such persons who need palliative care, there are a greater proportion of nervous diseases, circulatory diseases and genitourinary diseases.

The number of chronic diseases per pensioner under the age of 75 is less than for elderly people: in city it is 2.80 per man and 2.74 per woman (versus 2.96 and 2.93), in rural areas 1.38 and 1.28 (versus 1.58 and 1.47). In city, average number of chronic diseases per person aged 80 years and older is 2.89 (both among men and women), 2.86 for mobile patients (2.87 for men and 2.86 for women) and 3.25 for recumbent patients (3.60 and 3.20 respectively). In rural areas, this figure is 1.87 and 1.82 for men and women and 1.95 and 1.89 respectively for non-transportable elderly patients.

Averaged self-assessment of the health status rating on a four-point scale (1 - excellent, 4 bad) by pensioners under the age of 75 in city is more optimistic (2.6 versus 2.7) and in rural areas it is the same (2.8) as in the oldest age group. Average self-assessment of the health status in the older age group is 3.0 in city and 2.9 in rural areas.

Rural residents remain active longer than urban ones. Prevention of circulatory, digestive and genitourinary diseases probably promote the increase in life expectancy most of all. The structure of death causes does not adequately reflect the prevalence of chronic diseases and cannot

serve as a basis for studying the population's need for specialized medical care.

**Keywords:** health of elderly, burden of chronic diseases, living conditions, urban and rural residents

## Krissing Method with Fractional Distance Euclidean Matrices

**Leonidas Sakalauskas**

*Vilnius University, Akademijos st 4, Vilnius LT-08663, Lithuania,  
Shiauliai University, Vilniaus st 88, Shiauliai, LT-76285, Lithuania*

Euclidean distance matrices have been receiving increased attention because of their applications in recently active fields of research, such as molecular conformation in bioinformatics, dimensionality reduction in machine learning and statistics, semidefinite programming, wireless sensor network localization. The approach to kriging constructed in the paper, using Fractional Euclidean Distance Matrices (FEDM). The resulting model is rather simple and depends on small set of parameters (mean, variance and parameter of correlation function), which are efficiently estimated by the maximal likelihood method. Results of application of the FEDM model considered to analytically computed surfaces and mathematical modelling of a waste water filter design illustrated its applicability to scattered data analysis. The developed model allows prevision (values in the future) or reconstruction of missing data (values in the past). Of course, the model constructed might be generalised in multimodal case and noisy measurements.

## EM Algorithm for Estimating the Parameters of the Multivariate Stable Distribution

**Leonidas Sakalauskas<sup>1</sup>, Ingrida Vaiciulyte<sup>2</sup>**

*<sup>1</sup>Siauliai University, Lithuania <sup>2</sup>Siauliai State College, Lithuania*

Research of alpha-stable distributions is especially important nowadays, because they often occur in the analysis of financial data and information flows along computer networks. It has been found that financial data are often leptokurtic with a heavy-tailed distributions; many authors, e.g., Rachev, Mitnik (2000), Kabasinskas et al. (2012), Sakalauskas et al. (2013) have proved that the most often used normal distribution is not the most suitable way to analysis economic indicators and suggested to replace it with more general, for example, stable distributions. Since Rachev, Mitnik (2000), Kabasinskas et al. (2012), Sakalauskas et al.

(2013) have estimated one-dimensional alpha-stable distributions a problem arises how to estimate multidimensional data. Maximum likelihood method for the estimation of multivariate alpha-stable distributions by using EM algorithm is presented in this work. Integrals included in the expressions of the estimates have been calculated using the Gaussian and Gauss-Laguerre quadrature formulas. The constructed model can be used in stock market data analysis.

**Keywords:** Gaussian and alpha-stable model, EM algorithm, likelihood ratio test, quadrature formulas

## Choosing and Combining the Right Trees From an Ensemble

**Marco Sandri<sup>1</sup>, Roberto Savona<sup>2</sup>, Marika Vezzoli<sup>3</sup>**

<sup>1</sup> Data Methods and System Statistical Laboratory, University of Brescia, Brescia, Italy, <sup>2</sup> Dept of Economics and Management, University of Brescia, Brescia, Italy

<sup>3</sup> Dept of Molecular and Translational Medicine, University of Brescia, Brescia, Italy

In this paper we introduce a novel ensemble approach in the spirit of model clustering and combination. After generating a sample of bootstrap-based trees and identified their main clusters, the corresponding medoid trees are next combined by growing a tree on trees, namely a recursive partition using as covariates the predictions of the medoid trees. The resulting Final Classification Tree is then expected by construction to outperform single trees, being realized with the objective to maximize the forecasting power of single trees. Using data on 472 defaults and 471 non-defaults for non financial corporate in Italy during the period 2007-13, the research experiments prove that our synthetic model significantly outperforms all the representative trees also ranking among the best 1% distribution of the trees universe in terms of performance diagnostics.

**Keywords:** interpretability, Occam's razor problem, similarity between trees, Jaccard distance, corporate defaults.

## The Stochastic-Bayesian Connection within the Lévy Context

**Lino Sant**

Dept of Statistics and Operations Research, Faculty of Science, University of Malta, Malta

Through indexing of a suitably chosen collection of sets in tandem with a structure able to model sample path requirements, it has been proved

possible to generalize Lévy processes within extremely flexible constructs. These constructs have in turn enabled researchers to create Bayesian models of all sorts, especially hierarchical ones, with stables of docile prior probability families out of which posteriors can be computed. In this paper we concentrate on the link between Lévy processes which are set-indexed and corresponding Bayesian models while proposing novel ideas on how this can be exploited.

**Keywords:** Lévy processes, Bayesian models, priors.

## A Projection Pursuit Algorithm for Preference Data

**Mariangela Sciandra<sup>1</sup>, Antonio D'Ambrosio<sup>2</sup>, Antonella Plaia<sup>1</sup>**

<sup>1</sup>*Dept of Economics, Business and Statistics, Università degli Studi di Palermo, Italy,* <sup>2</sup>*Dept of Economics and Statistics, University of Naples Federico II, Italy*

In the framework of preference rankings, the interest can lie in finding which predictors and which interactions are able to explain the observed preference structures. The last years have seen a remarkable lowering of works about the use of decision tree for clustering preference vectors. As a matter of fact, decision trees are useful and intuitive, but they are very unstable: small perturbations bring big changes. This is the reason why it could be necessary to use more stable procedures in order to clustering ranking data. In this work, following the idea of Bolton (2003), a Projection Pursuit (PP) clustering algorithm for preference data will be proposed in order to extract useful information in a low-dimensional subspace by starting from a high but most empty dimensional space.

Projection pursuit clustering is a synthesis of projection pursuit and nonhierarchical clustering methods that simultaneously attempts to cluster the data and to find a low-dimensional representation of this cluster structure. As introduced by Huber (1985), a PP algorithm consists of two components: an index function  $I(\alpha)$  that measures the "usefulness" of projection and a search algorithm that varies the projection direction so as to find the optimal projections, given the index function  $I(\alpha)$  and the data set  $X$ . In this work a proper specified Projection index function for discrete data will be defined: several distances will be used to evaluate distances between the density of the projected data and the uninteresting uniform density. We also propose diagnostics for finding the optimum number of clusters in projection pursuit clustering. All the methodology is illustrated and evaluated on one simulated and one real dataset.

**Keywords:** Projection pursuit, preference data, Clustering rankings.

### References

1. R. Bolton and W. Krzanowski. Projection Pursuit Clustering for Exploratory Data Analysis. *Journal of Computational and Graphical Statistics*, 12, 1, 121{142, 2003.
2. P. J. Huber. Projection Pursuit. *Ann. Statist.*, 13, 2, 435{475, 1985.

## Development of Health Priorities for Elderly Residents of Russia and Belarus

**Victorya Semyonova and Tamara Sabgayda**

*Federal Research Institute for Health Organization and Informatics of Ministry of Health of Russian Federation*

In Russia, life expectancy increased by 7.6 years in males and 4.7 years in females since 2005. Similar processes are taking place in Belarus. The increasing share of older people in Russia and Belarus actualizes challenges, faced by Europe – health of older adults.

We analyzed development of health priorities for older adults aged 75+ in Russia and Belarus against the relevant European indicators using rates of European Union until May 2004 (EU-15) for 1989-2013 as references. As opposed to EU-15 where elderly mortality decreased almost linearly, in Russia and Belarus sustainable positive trends took shape only in 2005 resulting in an 18% mortality reduction in males and 15.6% in females in Russia in 1989-2013. In Belarus, in 1989-2013 the rates remained almost unchanged (mortality increased by 4.8% in males and decreased by 2.7% in females) versus an almost 30% reduction in EU-15.

The structure of elderly mortality in Russia and Belarus differs fundamentally from that of EU-15 due to abnormally high share of cardiovascular diseases against extremely low contribution of neoplasms: in 2013, it amounted to 64.5% and 59% in males and 70.4% and 55.1% in females versus 36.2% and 40.5% in EU-15. Contribution of neoplasms among the Russian and Belarusian males amounted to 13.1% and 9.3% versus 25.6 % in EU-15, in females - 8.5% and 5.7% versus 18.5%. Diseases of the respiratory system account for 5.2% of the total mortality in Russian and 2% in Belarusian males versus 11.7% in EU-15, and 2.1% and 0.4% versus 8.9% in females respectively. Even value of external causes in EU-15 turned out to be higher than in Russia and Belarus: 3.1% versus 2.4% and 2% in males and 2.8% versus 1.3% and 0.9% in females respectively.

Against this background, the share of ill-defined conditions in Russia and especially in Belarus turned out to be extremely high, equaling to 9.7% and 23.2% in males and 12.3% and 33.5% in females versus 3.3% and 4.5% in EU-15.

Mortality of older adults from ill-defined conditions is mainly formed by "senility" (R54). It is crucially important that, according to WHO criteria, "senility" (R54) is considered as unknown (unidentified) cause of death. However, nowadays ill-defined conditions are the third cause in older males and second in females in Russia, and second cause in Belarus regardless of gender. This is a very long situation developed by regulations adopted by the Soviet Union in 1989.

Concluding, we'd like to note that such situation is possible only if diagnosis among the elderly is very low and cardiovascular diseases are



listed as primary cause of death regardless of actual diagnosis. It leads to under-registered mortality from all other causes, including neoplasms, complicating development of health priorities for older adults in Russia and Belarus.

**Keywords:** mortality trends, accuracy of diagnosis of death causes, abnormally high share of cardiovascular diseases, extremely low contribution of neoplasms

## Exploring the Factors that Determine Depression among 50+ Europeans since Childhood: The Role of Adverse Experiences as Mediators

*Main topic: Health state of a population: definition, modeling and estimates*

**Eleni Serafetinidou<sup>1</sup> and Georgia Verropoulou<sup>1</sup>**

*Dept of Statistics and Insurance Science  
University of Piraeus, Greece*

Depression is a mental health condition causing major issues on individuals, families and society as a whole and affecting different periods of life. The present study aims to explore factors and predictors that may cause depression in later life in a population of 23288 Europeans aged 50 and higher in 2008. Observations were collected from the database of the Survey of Health and Retirement in Europe (SHARE). More specifically, the analysis includes persons participating both at wave 3 (SHARELIFE), which contains retrospective information from childhood and adulthood, and at wave 2 which contains information pertaining to the present. For the purposes of the analysis three successive binary logistic regression models were applied using SPSS, version 20. The first model isolates significant childhood factors predisposing to depression in later life. The second model identifies adulthood factors over the life course and the last model includes current factors. In the next step, differences between these models were examined. Finally, logistic regression models were applied for each gender separately, allowing for comparisons. Results showed that, for the total sample, childhood factors were significant predictors in model 1 and adulthood factors were significant in model 2 but going further, few of them didn't retained their significance in subsequent models. What is interesting is that all factors referring to present conditions in model 3 were significant in determining later life depression. Essentially, the last model identified the most important factors which may determine depression in older ages from childhood to the present time. The analysis of differences between genders in older ages is of special interest. Childhood factors are not as important for men while for women the opposite applies. Further, there is a differentiation in adverse experiences that can cause depression in later life for men and women. Finally, the

inclusion of predictors in model 3 proves that all of them are significant for both genders. Results will inform policy makers to focus on modifiable depression risk factors in order to take appropriate action.

**Key words:** gender, depression, adverse experiences, life course

**Acknowledgement:** *"This work has been completely supported by the General Secretariat for Research and Technology (GSRT) and the Hellenic Foundation for Research and Innovation (HFRI)".*

## Generalized Partially Linear Models to Predict the Goodness of a Child Garment Fit

Amelia Simó, M. Victoria Ibáñez, Irene Epifanio, Vicent Gimeno

*Dept of Mathematics – IMAC, Universitat Jaume I, Spain*

This work was motivated to study an important current application: size fitting for online children's garment shops.

To address this problem, the Biomechanics Institute of Valencia scanned a sample of Spanish children, getting each body shape represented by 3075 3D landmarks. Furthermore, a subsample of these children tested different garments of different sizes, and their fit was assessed by an expert as 'correct', 'small' or 'large'.

The aim of the work is to show how Generalized partially linear models can be used to predict the goodness of fit of a given garment size, i.e. small ( $Y_i = 1$ ), good fit ( $Y_i = 2$ ) or large ( $Y_i = 3$ ), as a function of the garment size, the size of the child and his/her shape (defined on the Kendall's Shape Space).

Partially linear models are regression models in which the response depends on some covariates linearly but on other covariates nonparametrically. They generalize standard linear regression techniques and are special cases of additive models, which makes it easier to interpret the effect of each variable.

Our aim is to define a generalized partially linear model (for an ordered ordinal response) on Riemannian manifolds (in particular on the Kendall's Shape Space), to develop and illustrate the algorithms for estimating it and to apply it to the children's garment fit problem.

**Keywords:** Shape space, Statistical shape analysis, Generalized linear models, Partially linear models, Kernel regression.

## Discretization of the Tail Density Function and Discrete Expected Percentage Shortfall

**George-Jason Siouris, Alex Karagrigoriou, Despina Skilogianni**

*Lab of Statistics and Data Analysis, Dept of Mathematics, University of the Aegean, Greece*

After extensive investigation on the statistical properties of financial returns, three properties have shown to be present in most, if not all, financial returns. Their existence has been the source of most problems associated with the estimation of the underlying risk of assets. These are often called the three stylized facts of financial returns and are volatility clusters, fat tails and nonlinear dependence.

In order to forecast the asset volatility, a number of different models have been developed over the years. Each of them offers an answer on a specific aspect of the problem at hand. Many of these models incorporate skewed, fat-tailed distributions. The disadvantage of this approach, is that even with the simple and well-known Student distribution (which has numerous ways of extending in order to support asymmetry) closed-form expected shortfall expressions are not available. This is also the case for many asymmetric heavy-tailed distributions.

A solution to this problem, is the discretization of the tail density function. This is a logical approach, since the nature of returns is discrete, as the market always operates on a specific accuracy. As a result of this discretization process, it is now possible to improve the expected shortfall estimations or, in our case, the expected percentage shortfall.

**Keywords:** Value at Risk, low price effect, low price correction, violation ration, backtesting

### References

- [1]. Siouris, George-Jason, and Alex Karagrigoriou. "A Low Price Correction for Improved Volatility Estimation and Forecasting." *Risks* 5.3 (2017): 45.

## The Health State of a Population and the related Curves: Historical Views and further Perspectives

**Christos H Skiadas<sup>1</sup>, Charilaos Skiadas<sup>2</sup> and George Matalliotakis<sup>3</sup>**

<sup>1</sup>*ManLab, Technical University of Crete, Chania, Crete, Greece*

<sup>2</sup>*Dept of Mathematics and Computer Science, Hanover College, IN, USA*

<sup>3</sup>*Hellenic Open University, Patra, Greece*

Following our recent publications in 2 books of "The Springer Series on Demographic Methods and Population Analysis" we provide a brief presentation of the main findings and improvements regarding the Health State of a Population and the related quantitative methods of estimation

based on the advanced stochastic theory of the first exit time. Several illustrative applications for various countries are presented.

The health state curve is illustrated in several graphs either as a fit curve to data or produced after a large number of stochastic realizations. The Health State, the Life Expectancy and the age at “mean zero health state” are also estimated.

**Keywords:** Health State and Survival Curves, Health status of a population, First exit time stochastic theory, stochastic simulations of health state, Age at Maximum Curvature, Healthy Life Expectancy and HALE, Standard Deviation  $\sigma$ , Health State Curves and other.

#### References

- [1]. Skiadas, C.H. and Skiadas, C. (2017). Exploring the Health State of a Population by Dynamic Modeling Methods. Springer, DOI 10.1007/978-3-319-65142-2.
- [2]. (see at: <https://link.springer.com/book/10.1007/978-3-319-65142-2> ).
- [3]. Skiadas, C.H. and Skiadas, C. (2018). Demography and Health Issues - 2018 Population Aging, Mortality and Data Analysis. Springer, ISBN 978-3-319-76002-5 (see at: <https://www.springer.com/gp/book/9783319760018#> ).

## Low Price Effect and Post Model Correction of Value at Risk

**Despina Skilogianni, Alex Karagrigoriou, George-Jason Siouris**

*Dept of Mathematics, University of the Aegean, Greece*

In finance, one of the main goals is the estimation of volatility, since it is crucial in risk analysis and management. In order to improve value at risk (VaR) forecasts, we discuss the concept of low price effect and introduce the low price correction which does not require any additional parameters and instead of returns it takes into account the prices of the asset. Judgment on the forecasting quality of the proposed methodology is based on a number of popular evaluation measures, which are properly adapted, in order to capture the needs of our problem. In addition, backtesting measures are also implemented. For illustrative purposes a real example from the Athens Stock Exchange is fully explored.

**Keywords:** Value at Risk, low price effect, low price correction, violation ration, backtesting

## Branching Processes Modeling in Cancer Research

**Maroussia Slavtchova-Bojkova**

*Sofia University "St. Kl. Ohridski", Bulgaria*

Stemming from the multistage theory of cancer, our main idea in this study is to introduce a novel multi-type decomposable branching process model

of a cell population, with  $n$  types  $n > 2$ . As it is stated in Durrett (2015) “Metastasis, the spread of cancer to distant organs, is the most common cause of death for cancer patients. It is very complex process: cell must enter the blood stream (intravasation), survive the trip through the circulatory system, leave the blood stream at its destination (extravasation), and survive in an alien environment, e.g., cells from the breast tissue living in bone”. Hence, we are exploring two main consequences from the biological nature of cancer: first, the metastatic process could involve more than one human organ and second, fallen in a completely different environment the cancerous cell may change its characteristics for living and division.

The present research is taking into account these facts in cancer development, namely, it is possible to have more than one type of metastasis in the human organism, possibly after local elimination of the initial tumor followed by proper medical treatment. In this sense, this paper comes as a continuation of the paper by Slavtchova–Bojkova et al. (2017) where a two-type branching process in continuous time is used to model mutations occurring in a population of successfully treated cancer cells.

**Keywords:** multi-type branching processes, probability of extinction, mutations, hazard function, heavy-tailed distributions.

**Acknowledgements.** *The research is supported by the National Fund for Scientific Research at the Ministry of Education and Science of Bulgaria, grant No DFNI-I02/17 and partially supported by and Ministerio de Economía y Competitividad and the FEDER through the Plan Nacional de Investigación Científica, Desarrollo e Innovación Tecnológica, grant MTM2015-70522-P, the Junta de Extremadura (grant GR15105), Spain.*

## References

- [1].Durrett, R. (2015). Branching Process Models of Cancer. Springer.
- [2].Slavtchova-Bojkova, M., Trayanov, P., Dimitrov, S. (2017). Branching processes in continuous time as models of mutations: Computational approaches and algorithms. Computational Statistics and Data Analysis,
- [3].<http://dx.doi.org/10.1016/j.csda.2016.12.013l>.

## Taylor Series Expansion Approach for Epistemic Uncertainty Propagation in Queueing Models with Inventory

**Soufit Massinissa, Abbas Karim**

*Research Unit LaMOS, Dept of Operational Research, Faculty of Exact Sciences, University of Bejaia, Algeria*

During the recent years there have been remarkable advances in modelling of complex systems whose performance measures are affected by numerous uncertain parameters. This has led to a growth in the area of uncertainty analysis as well as the need to develop robust and flexible

numerical methods for computing performance measures for such systems with uncertain inputs. The Taylor series expansions are now in widespread use for sensitivity analysis and uncertainty quantification. This paper addresses the problem of investigating if the sensitivity analysis and the uncertainty quantification can be integrated into same framework. More specifically, we provide an uncertainty analysis for queueing-inventory models, by extending the multivariate Taylor-series expansion methodology to such stochastic models, and then use Taylor-series expansion to compute Copulas theory. These theory is used to calculate the sensitivities of uncertain input parameters on the model output. We will also estimate probability density function and different moments of the output measure of the considered queueing-inventory model. The efficacy of the proposed method is shown with several numerical examples and obtained numerical results are compared to those of Monte Carlo simulation.

### **Probabilistic Models for Clinical Pathways: the Case of Chronic Patients**

**Stergiani Spyrou<sup>1,3</sup>, Anatoli Kazektsidou<sup>2</sup>, Panagiotis Bamidis<sup>1</sup>**

<sup>1,3</sup> *Lab of Medical Physics, Medical School, Aristotle University of Thessaloniki, Greece,* <sup>2</sup> *Region of Central Macedonia, Dept of Public Health and Social Welfare, Greece,* <sup>3</sup> *3<sup>rd</sup> Regional Health Authority (Macedonia) of Greece*

This paper presents the ongoing research on the use of Markov models on the electronic health records. The research aims to the reliability of the data flow and process flow of the different stages of the patient status following the health care processes in a health care setting. Specifically the use of mathematical modeling of the clinical pathways is examined through literature research. Clinical pathways translate best available evidence into practice, indicating the most widely applicable order of treatment interventions for particular treatment goals. A special focus will be given through a case study regarding the care of patients which are under the age of eighteen and have respiratory health problems. Applying models for the follow up of those patients and generally for chronic patients is innovative and of great importance. Special focus is given on chronic patients as the completion of the clinical pathway depends on parameters that implicitly relate to the provided health services such as the adulthood of children and other. The results of the models could become valuable knowledge tools to help the health care providers.

**Keywords:** Clinical Pathways, Markov models, Respiratory health problems

## On the Measurement of Positive Labour Mobility

**Maria Symeonaki, Glykeria Stamatopoulou**

*Dept of Social Policy, School of Social and Political Sciences,  
Panteion University of Social and Political Sciences, Greece*

The present paper aims at quantifying positive labour mobility for young individuals and studying its diversity across European countries. An index of positive labour mobility is introduced, in order to measure positive or desired transitions of young individuals among labour market states. Commonly used mobility indices that are well established in the literature take into account either the probability of staying in the same state or all transitions among labour market states. However, there is a question as to whether these kind of indices can represent a meaningful measurement of labour mobility, where we are more concerned with determining only the desired transitions, from unemployment or inactivity to employment, for example. This study proposes a new index which is also used in order to compare positive labour mobility of young individuals, aged between 15 and 29, among all European countries based on the latest at the time available raw data from the EU-Labour Force Survey (EU-LFS).

**Acknowledgement:** *This paper has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 649395 (NEGOTIATE – Negotiating early job-insecurity and labour market exclusion in Europe, Horizon 2020, Societal Challenge 6, H2020-YOUNG-SOCIETY-2014, Research and Innovation Action (RIA), Duration: 01 March 2015 – 28 February 2018).*

## Academic Success and Personal Characteristics Analysis by ANOVA Approach

**M. Filomena Teodoro<sup>1,2</sup>, Alcindo Gomes<sup>1</sup>, and Jose Maia Martins<sup>1</sup>**

*<sup>1</sup>CINAV, Escola Naval, Base Naval de Lisboa, Alfeite, 1900-001 Almada, Portugal, <sup>2</sup>CEMAT, Instituto Superior Tecnico, Universidade de Lisboa Av. Rovisco Pais, 1, 1046-01 Lisboa, Portugal*

The Naval Academy (NA) has a significant failure rate, losing about half of the admitted individuals during the course. This work aims to understand the causes associated to low income, through the analysis of the data of the individuals from the application to the end of the course, in order to identify the characteristics of individuals with greater and less likely to succeed in the NA. This work was started in [1]. In an initial phase, some techniques of descriptive statistics of data analysis were used. The first step of this analysis, the candidates are analyzed independently, and the admitted and the finalists are analyzed together comparing the variables at the beginning and at the end of the courses. Simple statistical

inference techniques were used [4], namely confidence intervals, parametric tests, contingency tables. Here we extend such analysis using intermediate level inference techniques, namely analysis of variance (ANOVA). We intend to complete this approach using a general linear models approach [2,3]. The study evidences greater success for individuals entering in NA with better grades and for individuals taking notice of the application competition over the internet.

**Keywords:** Indicators of academic success, Academic performance, Optimization of Education, Analysis of variance.

#### References

- [1]. A.D. Gomes, Rentabilizac~ao de Capital Humano, Master Thesis, Escola Naval, Almada, 2016.
- [2]. P. McCullagh, J.A. Nelder, Generalized Linear Models, Chapman & Hall, Londres, 1989.
- [3]. M.A. Turkman, G. Silva, Modelos Lineares Generalizados da teoria a pratica, Sociedade Portuguesa de Estatstica, Lisboa, 2000.
- [4]. A.C. Tamhane, D.D. Dunlop, Statistics and Data Analysis: from Elementary to Intermediate. Prentice Hall, New Jersey, 2000.

## A Constrained State Space Model with GARCH Effect for the Estimation of Hidden Asset Return Jumps

**Ourania Theodosiadou and George Tsaklidis**

*Dept of Mathematics, Aristotle University of Thessaloniki,  
54124 Thessaloniki, Greece*

In the present paper, a state space model is constructed in order to estimate the two-sided jumps of asset returns, which are unobservable. Trying to incorporate the phenomenon of conditional heteroskedasticity in the model, the noise terms of the state and measurement equations are assumed to have a time varying variance, that is modeled by a GARCH process. Then, the Kalman filter algorithm is used, along with the truncation of the jumps' probability density functions, in order to provide non negative estimations of the hidden components. Finally, we provide a change point estimation for the estimated hidden return jumps.

**Keywords:** Kalman filtering, GARCH effect, constrained optimization, asset return jumps.



## On Multiple Change-Point Estimation for Poisson Process

**Top, Alioune**

*University of Maine/France*

This work is devoted to the problem of change-point parameter estimation in the case of the presence of multiple changes in the intensity function of the Poisson process. It is supposed that the observations are independent inhomogeneous Poisson processes with the same intensity function and this intensity function has two jumps separated by a known quantity. The asymptotic behavior of the maximum likelihood and Bayesian estimators are described. It is shown that these estimators are consistent, have different limit distributions, the moments converge and that the Bayesian estimators are asymptotically efficient. Finally we present Monte-Carlo simulations to illustrate the obtained results and show that Bayesian estimator outperforms the MLE. Its concur also the i.i.d. case with one point of singularity (see [1] and [2]) where it was mentioned that the Bayesian estimators are generally more efficient than the MLE estimators in Change-Point type estimation problems.

The centerpiece of the method is the weak convergence of the normalized likelihood ratio process to an exponential functional of a two-sided difference of two Poisson processes driven by some parameters in a suitable metric space. In particular, we check the convergence of finite-dimensional distributions and the tightness of the corresponding family measures in the Skorohod space  $D_0(\mathbf{R})$ .

This result enables us to prove that our estimators of the jumps location converge with exact rate  $n^{-1}$  (better than in the regular case).

**Keywords:** Inhomogeneous Poisson process; change-point; Bayesian estimator; maximum likelihood estimator; likelihood ratio process:

### References

- [1]. Ibragimov, I. A., and Khasminskii, R. Z., (1981) Statistical Estimation. Asymptotic Theory, Springer, New York.
- [2]. Kutoyants, Yu. A., (1998) Statistical Inference for Spatial Poisson Processes, Lecture Notes in Statistics 134, Springer-Verlag, New York.

## Forecasting Population by Branching Processes

**Plamen Trayanov**

*Sofia University "St. Kliment Ohridski", Faculty of Mathematics and Informatics,  
Bulgaria*

In this paper the branching process theory is used to model human population and make forecasts about the future age structure and total population count. The Crump-Mode-Jagers branching process is used to

model the population of Bulgaria using data up to 2016. Simulations of the branching process are used to determine the confidence intervals of the forecast and then an analysis of the sources of risk to that forecast is done. As the laws of birth and death are not known ex-ante for the future, their uncertainty is contributing to the risk of the forecast. On the other hand, even if they are known in advance, the risk coming from the branching process itself must be evaluated. Finally, simulations of another crisis and a recovery are made, using historical data and their effect on the population is investigated.

**Keywords:** General Branching Process, Crump-Mode-Jagers Branching Process, population projections, demography

**Acknowledgements:** *The research was partially supported by the National Fund for Scientific Research at the Ministry of Education and Science of Bulgaria, grant No DFNI-I02/17.*

## Statistical Inference Based on $r$ -Size Biased Sample with an Application to Statistical Process Monitoring

George Tzavelas<sup>1</sup>, Apostolos Batsidis<sup>2</sup>, Polychronis Economou<sup>3</sup>

<sup>1</sup>*Dept of Statistics and Insurance Sciences, The University of Piraeus, Greece,*

<sup>2</sup>*Dept of Mathematics, University of Ioannina, Greece,*

<sup>3</sup>*Dept Civil Engineering, University of Patras, Greece*

Consistent and asymptotic normal estimators of the mean and the variance of a positive continuous random variable based on  $r$ -size biased samples under some non-restrictive conditions, are proposed. With the use of the above results univariate asymptotic control charts for monitoring the mean and the variance as well as multivariate control charts for the mean and the variance are constructed. Simulation results indicate the inefficiency of the asymptotic control charts for small and moderate sample sizes. To overcome this problem a parametric bootstrap approach is proposed which results to more efficient control charts.

**Keywords:** Biased sampling, Consistent estimators, Monitoring correction, Statistical process monitoring, Weighted distributions.

**Acknowledgements:** *This work has been partly supported by the University of Piraeus Research Center*

## **Forecasting Impact Factor of a Journal in Dentistry by PCA**

**P. Valderrama, M. Escabias, M.J. Valderrama**  
*Dept of Statistics, University of Granada, Spain*

From a stratified sample by tercils of journals in the field Dentistry, Oral Surgery and Medicine, a mixed linear regression model in principal components has been estimated in order to forecast the impact factor of any journal in a year starting from this information in previous years. Some additional explicative variables have also been considered.

**Keywords:** Principal component, impact factor, stepwise regression

## **Socioeconomic Inequalities in Depression and Cumulative Disadvantage Theory: Differences by Gender among Older Adults**

**Georgia Verropoulou, Eleni Serafetinidou**  
*University of Piraeus, Greece*

Childhood and adulthood socioeconomic status and adverse experiences have a significant effect on depression in later life. The main aim of the analysis is to assess the impact of socioeconomic inequalities, adverse experiences and poor health in childhood and over the life course in depression among older persons in the context of the cumulative disadvantage theory, which suggests that early life inequalities tend to increase over the life course. The analysis further considers differences between genders and across different age groups. The sample derives from combining cross-sectional data from the second wave of the Survey of Health, Ageing and Retirement in Europe (SHARE), carried out in 2006-2007, with retrospective information from SHARELIFE (wave 3), carried out in 2008-2009. The respondents included in the analysis are 23,816 persons, coming from 14 countries, spanning from South to North and from West to East: Greece, Germany, Sweden, Netherlands, Spain, Italy, France, Denmark, Austria, Switzerland, Belgium, Czech Republic, Poland and Ireland. For the purposes of the analysis binary logistic regression models were applied. The findings indicate that greater socioeconomic disadvantage, higher number of adverse experiences and more health problems are associated with higher chances of depression for both men and women and for all different age groups under consideration. However, the relative effect of cumulative disadvantage declines as age of the respondents increases, for both genders. Hence, it seems that accumulation of disadvantage leads to an increase in inequality but the relative effects decrease with increasing age.

**Keywords:** socioeconomic disadvantage, adverse experiences, life course, depression, cumulative disadvantage theory

## Models for Time Series whose Trend has Local Maximum and Minimum Values

**Norio Watanabe**  
Chuo University, Japan

Economical time series includes trend usually and it becomes important to capture trend adequately. Typical examples are series of stock prices or stock indices. In many cases trends show repeated up-and-down behavior. For such time series it is expected to predict time points of local maxima or minima and values themselves, and some prediction methods have been proposed. However, it is not clear whether those methods are appropriate or not. In this study we propose two kinds of models for time series whose trend has local maximum and minimum values. The first is a mean-stationary model. The second is a random walk type model. Proposed models provide basis for discussion about appropriateness of prediction methods. Simulation studies suggest that prediction is meaningless in some cases when time series has property of a random walk even though trend has local maximum and minimum values.

**Keywords:** stock index, random walk, prediction

## The impact of Definitions in Classifying the Employed, Unemployed and Inactive when Comparing Measurements from Different Sources

**Aggeliki Yfanti<sup>1</sup>, Catherine Michalopoulou<sup>1</sup>, Stelios Zachariou<sup>2</sup>**

<sup>1</sup>Dept of Social Policy, Panteion University of Social and Political Sciences, Greece, <sup>2</sup>Head of the Labour Force Division, Hellenic Statistical Authority, Greece

In all large-scale sample surveys and the census, gender, age, marital status, educational and occupational variables are included as background variables to provide information necessary for defining subpopulations and “contexts in which respondents’ opinions, attitudes, and behavior are socio-economically embedded”. Because these background variables play such a central role in social research, establishing their measurement’s cross-national and overtime comparability is essential and consequently, international classifications have been developed. However, in the case of the employment status, i.e. one of the occupational variables, all large-scale sample surveys and the

census use a perception question for its measurement whereas the European Union Labour Force Survey (EU-LFS) is using a synthesized economic construct according to the International Labour Organization (ILO) conventional definitions of the employed, unemployed and inactive. In this paper, we investigate the classification issues arising from these two different measurements. The analysis is based on the 2008-2014 EU-LFS annual datasets and the 2001 and 2011 Integrated Public Use Microdata Series (IPUMS)-International census datasets for Eastern (Hungary, Poland, Romania) and Southern (Greece, Portugal and Spain) Europe. The results are reported for the age group 15-74 so as to allow for comparability with the ILO conventional definition of unemployment.

**Keywords:** Employment status, ILO, EU-LFS, IPUMS-International, Eastern Europe, Southern Europe

## Distributions of some Run-Related Patterns in Random Sequences

**Yong Kong**

*School of Public Health, Yale University, USA*

Runs and pattern statistics have found successful applications in various fields. Many earlier results in the field were obtained by traditional methods, which are usually ad hoc, not easy to generalize, and when dealing with complicated patterns in multi-state or multiset systems, become intractable. Here we use a systematic approach that is inspired by methods in statistical physics [1]. In this approach the study of run and pattern distributions is decoupled into two easy independent steps. In the first step, elements of each object (usually represented by its generating function) are considered in isolation without regards of elements of the other objects. In the second step, formulas in matrix or explicit forms combine the results from the first step into a whole multi-object system with potential nearest neighbor interactions. By considering only one kind of object each time in the first step the complexity arising from the simultaneous interactions of elements from multiple objects is avoided. In essence the method builds up a higher level generating function for the whole system by using the lower level of generating functions from individual objects. Two examples, whose properties when the systems contain no repeated elements (symmetric permutation group) are well-studied, will be discussed in the context of arbitrary multisets (where each element can appear multiple times). The first example is the joint distributions of rises, falls, and runs [2]. The second example is the distributions of successions (also called 2-sequences) [3]. For both examples, explicit formulas for the generating function of the whole system, expectation, and (co)variance will be given. Potential generalization of the two-step method into a multiple-step method to handle more complicated patterns will also be discussed.

**Keywords:** Generating function, Permutations of multisets, Rises, falls, and levels, Runs statistics, Eulerian number and Simon Newcomb number, Successions.

#### References

- [1].Y. Kong. Distribution of runs and longest runs: A new generating function approach. *Journal of the American Statistical Association*, 101:1253–1263, 2006.
- [2].Y. Kong. Joint distribution of rises, falls, and number of runs in random sequences. *Communications in Statistics - Theory and Methods*, 2018a. doi:10.1080/03610926.2017.1414261.
- [3].Y. Kong. Distributions of successions of arbitrary multisets. manuscript, 2018b.

## Improved Predictive Modeling of Wood Composite Products using Bayesian Additive Regression Trees (BART)

**Timothy M. Young, Nana Tian, Yongke Sun**  
*University of Tennessee, USA*

This study presents real-time prediction models of modulus of rupture (MOR) and tensile strength from a wood composites manufacturing process. Variable preselection was used in the development of the various predictive models. Several regression models including multiple linear regression, partial least squares regression, neural networks, regression trees, boosted trees, bootstrap forest, and Bayesian additive regression trees (BART) were developed. BART had the best prediction performance in validation for both MOR and tensile strength. Specifically, validation results of MOR were promising with an average correlation coefficient across cross-validations of 0.86 and prediction error of 11.8%. Validation results for tensile strength had an average correlation coefficient across validations of 0.84 and 10%. The high prediction ability of BART is useful for manufacturers and researchers in improving the manufacturing process and reducing rework and reject losses; ultimately improving business competitiveness.

## Health Trends in the Entire and the Peripheral Populations of Greece during the 21st Century: a New Approach

**Konstantinos N. Zafeiris<sup>1</sup>, Christos Skiadas<sup>2</sup>**

<sup>1</sup>Laboratory of P. Anthropology, Dept of History and Ethnology, Democritus University of Thrace, Greece, <sup>2</sup>Manlab, Technical University of Crete, Greece

During the last decades life table analysis has proven to be one of the best and most parsimonious ways to estimate health levels in a population. In

this paper, 4 methods of estimating Healthy Years Lost in a population are presented and the average of the estimations given by them is proposed to be the most adequate measurement of Health levels. Namely, two of these methods are based on the Gompertz and Weibull fitting methods and the other two on the  $m(x)$  and  $q(x)$  distributions of a life table. These methods are illustrated in the Greek male and female population of the entire country. Also, the peripheral populations of Northern Greece, Central Greece, Attica and Aegean Islands (including also Crete) are used in the analysis in order to locate any differences found among them.

**Keywords:** mortality, health, life tables, Gompertz, Weibull

## **Mortality Differentials among the Euro-Zone Countries in 2016**

**Konstantinos N. Zafeiris<sup>1</sup>, Aspasia Tsoni<sup>2</sup>**

*<sup>1</sup>Laboratory of P. Anthropology, Dept of History and Ethnology, Democritus University of Thrace, Greece, <sup>2</sup>Dept of History and Ethnology, Democritus University of Thrace, Greece*

Euro-zone is consisting by 19 countries with different developmental and economic characteristics. During the first years of the 21<sup>st</sup> century several of them underwent a rapid economic and social crisis while others, like Germany, exhibited steadily economic and social growth. The scope of this paper is to compare the mortality experience of these countries using the last available data which correspond to the year 2016. The method used is a combination of a modified Heligman-Pollard procedure in addition with three cubic splines for the smoothing of the  $q(x)$  life tables distribution and the estimation of several parameters. Among them is life expectancy at birth, median and modal age at death, Kannisto's indicators and others. Results indicate the existence of significant differences among the countries studied.

**Keywords:** mortality, life tables, Heligman-Pollard, cubic splines

## **Empirical Study of Multivariate Tail Dependence**

**Ilze Zariņa, Gaida Pettere**

*Riga Technical University, Latvia*

Tail dependence is an important property of a copula. Bivariate tail dependence is investigated in many papers, but multivariate tail dependence for skewed data has not been studied widely. Skewed copulas are often used not only in financial problems but in insurance too. Loss distributions in insurance are usually heavily skewed. Therefore we

are going to study multivariate tail dependence for skewed data using recently introduced multivariate tail dependence coefficients.

**Keywords:** Skew t-copula, multivariate tail dependence coefficient, t-distribution, t-copula.



## Author Index

---

### A

Aars Jon · 37  
 Abbas Karim · 107  
 Abbruzzo Antonino · 17  
 Abdel-Aty Mahmoud · 13  
 Abdesselam Rafik · 13  
 Abola Benard · 14, 24  
 Afonso Filipe · 69  
 Aider Méziane · 66  
 Albano Giuseppina · 15  
 Amarioarei Alexandru · 94  
 Anderson Jane · 60  
 Andreopoulos Panagiotis · 15, 18  
 Antonietti Alessandro · 25  
 Antonov Valery · 16  
 Apsemidis Anastasios · 17  
 Asli Larbi · 66  
 Augugliaro Luigi · 17  
 Avraam Demetris · 18  
 Ayala Guillermo · 19

---

### B

Balakrishnan Narayanaswamy · 3  
 Balzan Maria · 19  
 Bamidis Panagiotis · 108  
 Banerjee Adrita · 20  
 Baran Sándor · 21  
 Barbu Vlad Stefan · 22  
 Bassanini Gisella · 82  
 Batsidis Apostolos · 112  
 Belaide Karima · 55  
 Bersimis Frangkiskos · 15, 18, 22  
 Bersimis Sotiris · 23, 44, 94  
 Bertarelli Gaia · 23  
 Biganda Pitos Seleka · 14, 24

Bongini Paola A. · 25  
 Boulougari Andromachi · 71  
 Bozeman James R. · 26  
 Brettschneider, J. · 34  
 Brown Mark · 3  
 Bushinskaya A.V. · 26, 27  
 Butt Asim · 28

---

### C

Caeiro Frederico · 29, 76  
 Cairns Karen J. · 77  
 Canhanga Betuel · 29, 72  
 Cappelli Carmela · 30  
 Carbone Anna · 92  
 Cardoso Debora · 54  
 Carette Philippe · 31  
 Cariou Véronique · 68  
 Carvalho João Batista · 32  
 Cazzaro Manuela · 82  
 Cera Fia · 87  
 Charalampi Anastasia · 33  
 Chia Ngee Choon · 33  
 Christodoulou, M. · 34  
 Cohen Samuel N. · 4  
 Coppola Mariarosaria · 35  
 Crippa Franca · 23, 36  
 Cubaynes Sarah · 37  
 Cubaynes Sarah · 95

---

### D

D'Amico Guglielmo · 38, 40, 41  
 D'Ambrosio Antonio · 101  
 de Brun Ciara · 74  
 De Gaetano Davide · 15  
 di Iorio Francesca · 30  
 di Prignano Ernesto Volpe · 38

Diday Edwin · 69  
Dimotikalis Yiannis · 41  
Domingues Edson · 54  
Doquenne Marie-Noelle · 42  
Douli Marigo · 43

---

## **E**

Economou Polychronis · 44, 112  
Edeleva Anna · 98  
Elliott Robert J. · 4  
Engström Christopher · 14, 24  
Epifanio Irene · 104  
Escabias M. · 113

---

## **F**

Farha Sultana · 44  
Farmakioti Evanthia · 88  
Farmakis Nikolaos · 45, 88, 89  
Fazekas István · 45  
Fotopoulos Stergios B. · 46  
Freeman Jim · 47  
Fujii Takuya · 47

---

## **G**

Gaillard Jean-Michel · 48, 97  
García, Jesús E. · 48, 49  
Giannouli Panagiota · 49  
Gimenez Olivier · 37  
Gimeno Vicent · 104  
Gismondi Fulvio · 38, 40, 41  
Gomes Alcindo · 109  
González-López, V.A. · 48, 49  
Grabski Franciszek · 50  
Griffin Larry D. · 50  
Grilo Luís M. · 51  
Groll Andreas · 52

Grossi Luigi · 53  
Guedes Gilvan · 54  
Guerry Marie-Anne · 31, 96  
Guillen Montserrat · 38

---

## **H**

Haddad Soraya · 55  
Hadjiliadis Olympia · 4  
Hölzl W. · 55

---

## **I**

Iannello Paola · 25  
Ibáñez M. Victoria · 104

---

## **J**

Jandhyala Venkata K. · 46  
Janssen Jacques · 38

---

## **K**

Kakava Konstantina · 56  
Kaklamani Stamatina · 42  
Kakuba Godwin A. · 14, 24, 70, 81  
Kalamaras Dimitrios · 56  
Kalamatianou Aglaia · 56, 75  
Kalligeris Emmanouil-Nektarios · 57  
Kang Yang · 58  
Kaniovski S. · 55  
Kaniovski Y.M. · 55  
Karadima Mary · 78  
Karagrigoriou Alex · 22, 57, 90, 105, 106  
Kazektsidou Anatoli · 108  
Keikara Muhumuza Asaph · 70

**Khan Naveed** · 58  
**Khokhlov Yury** · 59  
**Kimmel Barbara** · 60  
**Kokkinioti Theodora K.** · 64  
**Kollo Tõnu** · 61  
**Konstantinides Dimitrios G.** · 62, 63  
**Korenjak-Černe Simona** · 69  
**Kostaki Anastasia** · 63  
**Kotzamanis Byron** · 63  
**Kountzakis Christos E.** · 49, 63  
**Koutras Markos V.** · 5, 23, 64  
**Koutras Vasileios M.** · 65  
**Koutras Vasilis P.** · 66, 95  
**Kovalenko Anatoly** · 16  
**Krska Janet** · 73  
**Kwak Minsuk** · 91

---

## L

**Labenne Amaury** · 68  
**Lambrou Angeliki** · 57, 90  
**Lebedinskii Konstantin** · 16  
**Lee Tae Rim** · 67  
**Lefèvre Claude** · 6  
**Lemaitre Jean-François** · 48, 97  
**Li Nan** · 68  
**Limnios N.** · 6  
**Llobell Fabien** · 68  
**Loh Shu Peng** · 33  
**Londoño, Sérgio** · 48  
**Losidis Sotirios** · 69  
**Lotrič Dolinar Aleša** · 69  
**Loumponias Kostas** · 70  
**Lundengård Karl** · 70, 71, 81

---

## M

**Magalhaes Aline** · 54  
**Magero John Mango** · 24  
**Majewska Justyna** · 72

**Makrides Andreas** · 22  
**Malefaki Sonia** · 66  
**Malyarenko Anatoliy** · 29, 72  
**Manca Raimondo** · 2, 38  
**Manfrin Andrea** · 73  
**Mango John** · 14, 70, 81  
**Mar Molinero Cecilio** · 73  
**Mariani Paolo** · 74  
**Marletta Andrea** · 74  
**Marshall Adele H.** · 74, 75, 77  
**Martin Donald E.K.** · 75  
**Martins Jose Maia** · 109  
**Masala Giovanni** · 38  
**Masserini Lucio** · 74  
**Matalliotakis George** · 105  
**Mateus Ayana** · 76  
**Mazzoleni Marcella** · 76  
**McClellan Sally** · 58  
**McCurdy Meabh G.** · 77  
**Mecatti Fulvia** · 23  
**Mi Hong** · 68  
**Michalelis Georgios I.** · 78  
**Michalopoulou Catherine** · 33, 114  
**Minkova Leda D.** · 79  
**Mitchell Hannah** · 79  
**Mohri Hiroaki** · 80  
**Mori Jillian** · 26  
**Muhumuza Asaph Keikara** · 81  
**Mukhopadhyay Barun Kumar** · 81

---

## N

**Nan Fany** · 53  
**Negri Ilia** · 82  
**Neves Manuela** · 93  
**Ni Ying** · 29, 72  
**Nicolussi Federica** · 82  
**Nishihara Michi** · 83  
**Noe Laurent** · 75  
**Noronha Kenya** · 54  
**Nugent Chris** · 58

---

**O**

Oliveira Amílcar · 84, 85  
 Oliveira Manuela · 29  
 Oliveira Teresa · 51, 84, 85  
 Österberg Jonas · 70, 81

---

**P**

Paes Neir Antunes · 32  
 Palikrousis Haris · 88  
 Panagiotakos Demosthenes · 22, 86  
 Panaretos Dimitris · 86  
 Papachristos Apostolos · 87  
 Papadimitriou Elephtheria · 71  
 Papadopoulou Aleka · 88  
 Papageorgiou Myrto · 88  
 Paparas Alex · 46  
 Papatsouma Ioanna · 89  
 Parpoula Christina · 57, 90  
 Pecsora Sándor · 45  
 Peköz Erol · 7  
 Petar Jevtic · 91  
 Petroni Filippo · 38, 40, 41  
 Pettere Gaida · 117  
 Pirvu Traian · 91  
 Plaia Antonella · 101  
 Platis Agapios N. · 66, 95  
 Politis Konstadinos · 69  
 Ponta Linda · 92  
 Potrykowska Alina · 92  
 Pradel Roger · 37  
 Prata Gomes Dora · 93  
 Preda Cristian · 94  
 Psarakis Stelios · 17, 94  
 Psomas Panagiotis · 95

---

**Q**

Qannari El Mostafa · 68

---

**R**

Rancic Milica · 29, 71  
 Richardson Clive · 33  
 Riffo-Campos Angela · 19  
 Rinaldi Emanuela E. · 25  
 Robine Jean-Marie · 7, 95  
 Rodi Lykou · 71  
 Rombaut Evy · 96  
 Ronget Victor · 97  
 Rosaria Simone · 30, 35  
 Russolillo Maria · 35

---

**S**

Sabgayda Tamara · 98, 102  
 Sachlas Athanasios · 23, 44, 94  
 Sakalauskas Leonidas · 99  
 Salvadore Abigail · 26  
 Sambt Jože · 69  
 Sandri Marco · 100  
 Sant Lino · 19, 100  
 Santiago Flaviane · 54  
 Savona Roberto · 100  
 Schauburger Gunther · 52  
 Sciandra Mariangela · 101  
 Seirinopoulos Dimitrios · 45  
 Semyonova Victorya · 98, 102  
 Serafetinidou Eleni · 103, 113  
 Shegog Ross · 60  
 Shiohama Takayuki · 47  
 Sikolya Kinga · 21  
 Silva Domingos · 29  
 Silvestrov Dmitrii · 8  
 Silvestrov Sergei · 14, 24, 29, 70, 71, 72, 81

**Simó Amelia** · 19, 104  
**Singh Ajeet Kumar** · 20  
**Singpurwalla Nozer D.** · 9  
**Siouris George-Jason** · 105, 106  
**Skiadas Charilaos** · 105  
**Skiadas Christos H** · 105, 116  
**Skilogianni Despina** · 105, 106  
**Slavtchova-Bojkova Maroussia** ·  
 106  
**Soufit Massinissa** · 107  
**Souza Kenia** · 54  
**Spyrou Stergiani** · 108  
**Stamatopoulou Glykeria** · 109  
**Steinsaltz, D.** · 34  
**Strass Belinda** · 71  
**Sun Yongke** · 116  
**Swank Paul** · 60  
**Swishchuk Anatoliy** · 10  
**Symeonaki Maria** · 109

---

## T

**Tagini Angela** · 36  
**Takeshita Jun'ichi** · 80  
**Teodoro M. Filomena** · 109  
**Theodosiadou Ourania** · 110  
**Tian Nana** · 116  
**Timashev S.A.** · 26, 27  
**Top, Alioune** · 111  
**Tragaki Alexandra** · 15  
**Tran Tuan Q.** · 91  
**Trayanov Plamen** · 111  
**Triantafyllou Ioannis S.** · 64  
**Trzpiot Grażyna** · 72  
**Tsaklidis George** · 70, 71, 110  
**Tsoni Aspasia** · 117  
**Tzavelas George** · 86, 112

---

## V

**Vaiciulyte Ingrida** · 99  
**Valderrama M.J.** · 113  
**Valderrama P.** · 113  
**Valge Marju** · 61  
**Vamvakari Malvina** · 22, 86  
**Varlamis Iraklis** · 22  
**Vassiliou P.-C.G.** · 10  
**Verropoulou Georgia** · 87, 103, 113  
**Vezzoli Marika** · 100  
**Vigneau Evelyne** · 68  
**Vinciotti Veronica** · 17

---

## W

**Watanabe Norio** · 114  
**White Chris** · 28

---

## Y

**Yarovaya Elena** · 11  
**Yfanti Aggeliki** · 114  
**Yoccoz Nigel** · 37  
**Yogesh Mani Tripathi** · 44  
**Yong Kong** · 115  
**Young Timothy M.** · 116

---

## Z

**Zachariou Stelios** · 114  
**Zafeiris Konstantinos** · 63, 116, 117  
**Zariņa Ilze** · 117  
**Zenga Mariangela** · 25, 74, 75  
**Zhang Shuai** · 58